MACHINE LEARNING ENABLES THE USE OF SPECTROSCOPIC MRI TO GUIDE RADIATION THERAPY IN PATIENTS WITH GLIOBLASTOMA

A Dissertation
Presented to
The Academic Faculty

by

Saumya Suresh Gurbani

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Wallace H. Coulter Department of Biomedical Engineering

Emory University School of Medicine
Georgia Institute of Technology College of Engineering
March 2019

MACHINE LEARNING ENABLES THE USE OF SPECTROSCOPIC MRI TO GUIDE RADIATION THERAPY IN PATIENTS WITH GLIOBLASTOMA

Approved by:

Dr. Hyunsuk Shim, Advisor
Department of Radiation Oncology
Emory University School of Medicine
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Lee Cooper, Advisor
Department of Biomedical Informatics
Emory University School of Medicine
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Melissa Kemp
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Peng Qiu
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Hui-Kuo Shu
Department of Radiation Oncology
Emory University School of Medicine

Dr. David Yu
Department of Radiation Oncology
Emory University School of Medicine

Date Approved: March 13, 2019

ACKNOWLEDGEMENTS

The work in this dissertation could not have been completed without the support of many members of an interdisciplinary and interinstitutional team. First and foremost, thanks go to my co-advisors, Drs. Hyunsuk Shim and Lee Cooper, who have guided me throughout my doctoral research and whose continued support I look forward to throughout my career. Next, the many members of our research team at Emory: Dr. J. Scott Cordova, Dr. Hui-Kuo Shu, Dr. Brent Weinberg, Dr. Eduard Schreibmann, Dr. Jeffrey Olson, Mr. Michael Larche, Mrs. Sarah Basadre, Mr. Robert Smith, Ms. Samira Yeboah, Mr. Karthik Ramesh, Dr. Zhongxing Liang, Dr. Eric Salgado, Dr. Younghoon Yoom, Dr. Youn Oum, Dr. Lei Zhou, Mr. Pooya Mobadersany, Ms. Safoora Yousefi, and Dr. Mohamed Amgad; and to our collaborators: Dr. Andrew Maudsley, Mr. Sulaiman Sheriff, Dr. Eric Mellon, Dr. Peter Barker, Dr. Lawrence Kleinberg, and Dr. Sineyob Ahn. Finally, thanks go to the members of my dissertation committee heretofore not mentioned for their guidance in preparing this work: Dr. Melissa Kemp, Dr. Peng Qiu, and Dr. David Yu.

I'd also like to thank the Emory University Medical Scientist Training Program and its leadership: Dr. Robert Gross, Dr. Mary Horton, Dr. Cathy Quinones-Maeso, and Ms. Erica Weaver.

I'd like to acknowledge my family and friends who have supported me throughout the past four years. To my parents, who have inspired me throughout my life; to my brother and sisters; to my fiancée and best friend, Sarah Dupont; and to my friends who have been there through thick and thin.

Finally, and importantly, I'd like to thank the patients who have graciously agreed to be a part of this research; they are the reason we pursue this research, so that we may one day improve the lives of future patients.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AUC	area under the curve
CEMD	convolutional encoder – model decoder
Cho	choline
CNN	convolutional neural network
Cr	creatine
DSC	Dice similarity coefficient
EPSI	echo planar spectroscopic imaging
FLAIR	fluid-attenuation inversion recovery
GPU	graphical processing unit
MIDAS	Metabolite Imaging and Data Analysis System
MRS	magnetic resonance spectroscopy
MRSI	magnetic resonance spectroscopic imaging
NAA	N-acetylaspartate
ppm	parts per million
RMSE	root-mean-squared error
ROC	receiver-operator characteristic
RT	radiation therapy
sMRI	spectroscopic magnetic resonance imaging
T	Tesla
WHO	World Health Organization
ϕ_0	zero-order phase shift
ϕ_1	first-order phase shift

SUMMARY

Glioblastoma is the most common adult primary brain tumor and is highly aggressive due to its diffusely infiltrative nature. Radiation therapy has been shown to be the best single treatment for improving prognosis but requires accurate pre-therapy imaging for proper radiation planning. Spectroscopic magnetic resonance imaging (sMRI) is an advanced imaging modality that measures specific in vivo metabolite levels within the brain and has shown to be highly sensitive and specific in the detection of proliferative pathology. Clinical application of sMRI has been extremely limited due to computational challenges in sMRI data analysis. In this work, we utilize novel machine learning architectures to develop a software framework to close the gap for clinical utilization of sMRI in radiation therapy planning. First, we develop convolutional neural network to identify and remove spectral artifacts that lead to erroneous measurement. Next, we develop an algorithm for internally normalizing sMRI volumes, enabling voxel-to-voxel comparison across subjects and allowing threshold-based techniques to be used for target delineation. Third, we create a novel unsupervised learning framework to perform accelerated spectral quantitation, reducing the computational time and power needed to utilize sMRI. Finally, we develop a web-based software framework that bridges the gap between sMRI and its clinical use and demonstrate the feasibility of using this software in a multi-site clinical study to guide a radiation boost to regions of metabolic abnormality in patients with glioblastoma.

INTRODUCTION

As advances in precision medicine continue, medical imaging will serve as a central component in the diagnosis, monitoring, and prognostication of diseases. Medical imaging enables physicians to gain insight into a disease state and provide supplemental information to the clinical workup. To quote a 2003 opinion paper in the American Journal of Roentgenology: "To strike a target, it is necessary to see it. […] In what context does the contemporary physician most frequently encounter the internal anatomy of the human body? […] For most physicians, […] the answer is radiology." (1) The necessity of medical imaging is highly evident in the clinical workflow for patients with brain cancers, where imaging comes into play from the time of diagnosis throughout treatment and follow-up. In this work, we focus specifically on the use of an advanced imaging technique, spectroscopic magnetic resonance imaging, for patients with glioblastoma.

This chapter begins with literature reviews of the standard-of-care therapy for glioblastoma, of spectroscopic magnetic resonance imaging, and of the use of machine learning in the medical imaging field. It concludes with a presentation of the specific aims of this dissertation.

1.1 Therapy for Glioblastoma

1.1.1 Epidemiology

Glioblastoma, previous referred to as glioblastoma multiforme or GBM, is a highly aggressive primary brain tumor of glial cells. Glial cells are the neuron-supporting cells of the brain, playing a variety of roles in maintenance of neuronal health, including: myelination of neuronal axons, formation of the blood-brain barrier, regulation of the neuronal environment, and maintenance of necessary levels of blood-borne nutrients and oxygen to neurons (2). There are three main classes of glia in the central nervous system: microglia, astrocytes, and oligodendrocytes. Molecular mutations and pathway dysregulations in each of these types can lead to tumors, collectively known as gliomas. Historically, gliomas were classified by cell-type and their appearance on histological pathology. Recent work by consortia such as The Cancer Genome Atlas led in 2016 to new classifications by the World Health Organization (WHO) for gliomas based on molecular subtype and genetic mutations (3). This codification enables a better understanding of the underlying causes of these tumors and the progression of disease from low-grade gliomas to high-grade gliomas. Glioblastoma is the WHO Grade IV glioma, formed from unchecked proliferation of astrocytes. Glioblastomas are the most common amongst the malignant primary adult brain tumors, accounting for 15% of all neuroepithelial malignancies, with an incidence in the United States of 3.2 per 100,000 persons (4).

The standard of care for patients with glioblastoma is an aggressive regimen consisting of surgery, radiation therapy (RT), and chemotherapy (5). First, patients will undergo maximal safe surgical resection (debulking) of the tumor. Next, high dose radiation will be delivered to the surgical cavity and any residual tumor. For both surgery and RT, medical imaging is necessary to define targets for treatment. Concurrent and adjuvant chemotherapy with temozolomide, a DNA alkylating agent and the first-line

chemotherapy for glioblastoma, is continued throughout the patient's disease course. However, despite these aggressive treatments, median survival remains poor at just 15 months post-diagnosis (5). There are several hypotheses as to why this regimen still leads to poor outcomes, one of which is based on the presence of tumor that is undetected by standard medical imaging protocols and is discussed in the next section.

1.1.2 The role of imaging in glioblastoma

Magnetic resonance imaging (MRI) is the imaging modality used for diagnosing and planning treatment for patients with glioblastoma. MRI relies on the quantum spin of charged particles (e.g. protons) and how this spin can be manipulated in a magnetic field to yield a detectable imaging signal that can differentiate tissues of interest. Two different MRI techniques are commonly used which provide complementary sensitivity and specificity of tumor detection.

High dose RT is targeted using a technique known as contrast-enhanced T1-weighted magnetic resonance imaging (CE-T1w MRI). In this imaging modality, an external contrast agent, typically gadolinium, in injected intravenously into the patient. In the healthy brain, the gadolinium from the systemic blood supply will not be able to enter the brain because of the intact blood-brain barrier. The blood-brain barrier is composed of astrocyte foot processes and the endothelial cells of cerebral blood vessels, which together form a barrier that limits the free diffusion of large or highly charged particles across the endothelial cell membrane (6). In glioblastoma, however, the blood-brain barrier breaks down in the most proliferative sections of tumor, allowing compounds such as gadolinium to leak into the

brain parenchyma. In CE-T1w MRI, the T1 signal – a measurement of the early relaxation of protons in a magnetic field after a radiofrequency pulse, also known as spin-lattice relaxation - is used to define the image intensity for every voxel in the brain (7). Since gadolinium has a much higher T1 relaxivity compared to endogenous tissue, any gadolinium in the parenchyma will show up as bright enhancement on CE-T1w and can be used to detect regions of tumor with high specificity (8). A dose of 60 Gy is delivered to any residual enhancing tissue (after debulking of tumor), since these regions are known to be of high tumor burden. However, since gadolinium will only leak into to the tissue where the blood-brain barrier has been disrupted, CE-T1w MRI does not identify regions of tumor where the barrier is intact. Thus, a more sensitive imaging technique is used to guide a second prescription dose of radiation.

A second tier of RT (typically 46-51 Gy) is delivered to areas of hyperintensity on T2-weighted fluid-attenuation inversion recovery (FLAIR) MRI (9). The T2-weighting of signal, also known as spin-spin relaxivity, is a measurement of the later relaxation of protons in a magnetic field after radiofrequency stimulation (7). A variety of possible tissue pathologies can lead to increases in the T2-weighted signal as measured by FLAIR: tumor, inflammation, or vasogenic edema (10). Thus, FLAIR provides a sensitive but not specific measure of tumor burden.

Despite this high dose of RT and concurrent chemotherapy, patients often undergo disease progression (recurrence) at a median of 4 months after treatment (5). Recurrent glioblastoma is very difficult to treat, often being resistant to further radiation and inaccessible for secondary surgical resection (11); therefore, understanding why tumor recurs and trying to delay recurrence becomes a primary goal of glioblastoma research.

The location of recurrent disease can also vary largely: within the original 60 Gy RT target, within the intermediate dose area, or to regions several centimeters away, including crossing the midline (12). The latter of these suggests that there were regions of occult tumor which were not targeted for RT, and therefore able to continue proliferating. It is therefore necessary to find a better imaging modality that can identify the full extent of tumor so that these regions of occult tumor burden can be appropriately targeted for treatment. After all, what cannot be seen cannot be treated.

1.2 Magnetic resonance spectroscopy

Proton magnetic resonance spectroscopy (^{1}H-MRS) is a modality that enables the *in vivo* measurement of several different chemicals, or metabolites, in tissue. Unlike proton MRI, which measures the bulk resonance signal of all protons in a magnetic field, MRS separates out the signals of protons from different chemicals. While protons have a specific resonance frequency in a magnetic field, 42.58 MHz / T, local electric fields around each proton due to the presence of electrons can slightly shift the resonance frequency on the order of tens of Hz. A proton that is in a relatively neutrally-charged environment, such as within the hydrocarbon chain of a lipid, will have a resonance frequency shift close to zero. n the other hand, protons next to lots of electrons, such as the two in each molecule of water, will have a larger shift. This chemical shift is described in the number parts per million (ppm), indicating the shift in Hz per MHz from the central resonance frequency; in this way, this unitless measure of chemical shift enables comparison across different field strengths. MRS can be performed on standard MR instrumentation, and a single-voxel implementation is available from most vendors. Figure 0.1 shows a sample single-voxel

MRS screenshot from a vendor-provided sequence. As shown, several cerebral metabolites can be detected using short-to-medium echo times at acquisition (13-15):

- Choline-containing compounds (Cho, 3.2ppm) are associated with phospholipid metabolism and membrane synthesis; the Cho signal can be used as a surrogate measure of membrane turnover in proliferating cells;
- Creatine (Cr, 3.0ppm and 3.9ppm) is an energy-carrying compound associated with ATP synthesis;
- N-acetylaspartate (NAA, 2.0ppm and 4.4ppm) is a neuronal protein used in energy transfer and myelin metabolism found specifically in neurons; it is therefore used as a measure of neuronal health.

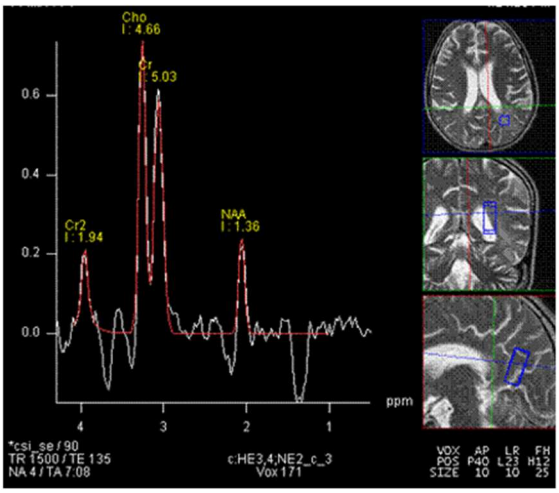


Figure 0.1: Sample single-voxel MRS of the brain from a vendor-provided sequence.

1.2.1 Whole-brain spectroscopic MRI identifies occult tumor

While MRS has been available for decades – predating MRI as we know it today (16) – the technology has only recently been adapted for imaging the whole-brain rather than measuring metabolite signal in a single voxel. The echo-planar spectroscopic imaging (EPSI) sequence created by Andrew Maudsley's group at the University of Miami is one example of a modern MRS imaging sequence (17,18). EPSI uses a pulse-echo paradigm to measure multiple voxels simultaneously in a single slice; multiple slices are measured in series to produce a 3D volume. EPSI incorporated an interleaved water measurement in addition to the chemical shift measurements, which enables the normalization of each metabolite level to the water signal, thus yielding metabolite concentrations. The current implementation of EPSI on a 3T scanner produces a 64x64x32 matrix with a spatial resolution of 4.5x4.5x5.4mm, with an acquisition time of 15 minutes. During the same scanning session, a high-resolution (~1mm isotropic) T1w sequence is obtained. The T1w and metabolite measurements can then be co-registered into the same image space, allowing for metabolite overlays on the structural information provided by T1w MRI (Figure 0.2). This combination of high-resolution whole-brain MRS aligned with conventional imaging has been dubbed "spectroscopic magnetic resonance imaging" (sMRI) by Shim et. al. (19,20).

sMRI has been shown to be useful in the detection of metabolite alterations in several brain pathologies, including tumors such as glioblastomas. In glioblastoma, the malignant glial cells proliferate rapidly and therefore have high membrane turnover; when compared to the relatively senescent neurons, this produces a detectable increase in choline metabolism; concurrently, the proliferation of glia leads to displacement and death of

neurons, resulting in a marked decrease in the levels of intracellular N-acetylaspartate (21,22). As shown in Figure 0.2, the regions of the brain with such altered metabolism can be mapped out using sMRI. Recently, our lab has shown that sMRI changes in these two metabolites and their ratio (Cho/NAA) can identify occult tumor beyond what is seen on standard imaging (20). In that work, Cordova et. al. scanned patients with glioblastoma using the EPSI sequence, then, at the time of resection, collected tissue samples from regions of elevated Cho/NAA. Some of these samples were within regions of enhancement on CE-T1w or hyperintensity on FLAIR imaging, but several were in regions that appeared "normal" on conventional imaging. However, all samples were found to contain at least some tumor cells in histology (as measured by staining with SOX2, a glioma-specific marker), and Cho/NAA was found to have a statistically significant correlation with the density of tumor cells on histology (19,20).

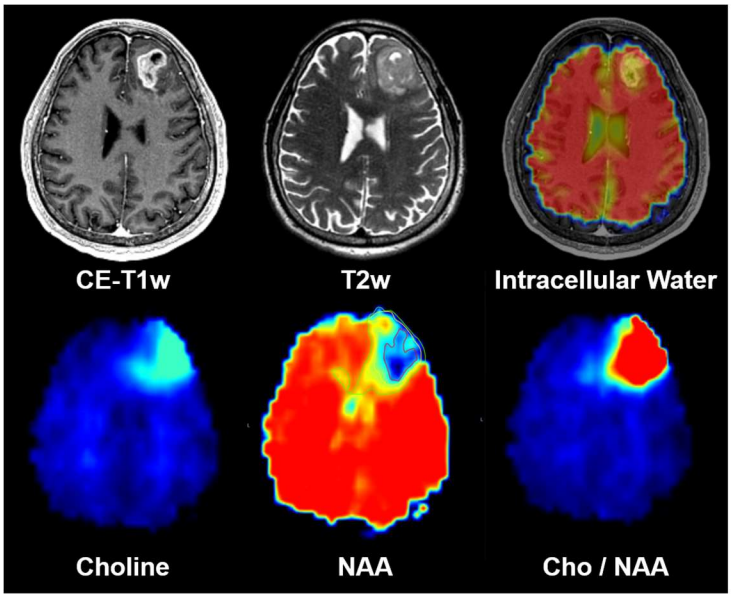


Figure 0.2: Spectroscopic MRI (sMRI) of a patient with glioblastoma.

Furthermore, Cordova et. al. showed that regions of Cho/NAA elevation that were occult on standard imaging – and thus not targeted for surgery or high-dose RT – correlated with the regions where disease first recurred. These landmark findings strongly suggest that sMRI can be a useful biomarker for identifying occult tumor, and could potentially be used to proactively modify where treatment is delivered to patients.

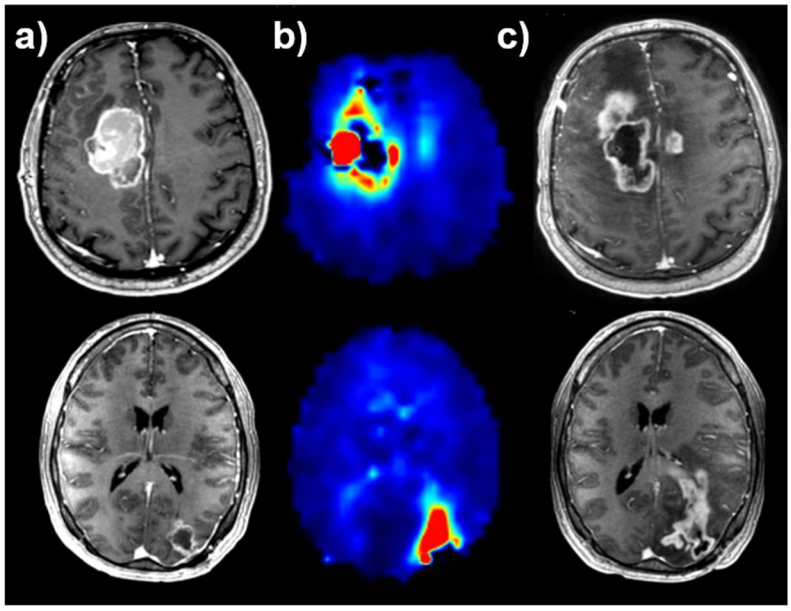


Figure 0.3: The Cho/NAA ratio can identify occult regions of GBM infiltration. Two cases are shown; (a) CE-T1w MRI taken pre-RT reveals enhancing tumor only, while (b) concurrent sMRI reveals elevated Cho/NAA abnormalities beyond enhancing lesions; (c) a CE-T1w MRI taken 4 weeks post-RT reveals tumor growth along both sMRI abnormalities. Adapted from Cordova et. al., Neuro-Oncology (2016).

1.3 Current Challenges for sMRI

While MRS has been shown to be useful in several neuropathologies, many technical challenges have prevented wide-spread adoption of this technique. While research is actively being pursued on improving MRS acquisitions, this section will focus on post-acquisition processing since that is most relevant to the work in this dissertation, and is of particular importance to whole-brain MRS.

1.3.1 Spectral Fitting

As with other MRI sequences, the signals collected from scanner consist of field induction decays (FIDs) as different parts of k-space are acquired. These FIDs are composed of several sinuosoidal components oscillating at the resonance frequencies of the protons they represent. Once these FIDs are acquired, they are converted into the frequency domain via Fourier transform as seen in Figure 0.1. The next step is to quantify the metabolites present within the spectrum by applying known models of spectral shape to the acquired data, and is often referred to as "spectral quantitation", "spectral analysis", or "spectral fitting".

A common model for describing the shape of proton resonance in a magnetic field is the Lorentzian-Gaussian function (23). In the time domain, the function describes a dampened oscillation at a given frequency, while in the frequency domain it describes a single resonance peak centered on the same frequency. There are six parameters for the Lorentzian-Gaussian function: frequency, amplitude, zero- and first-order phase, and two linewidth coefficients representing the combination of Lorentzian and Gaussian dampening. Variations on this model include using just a Gaussian or Lorentzian shape, or the Voigt profile (24). Algorithms which model a spectrum as a combination of individual peak resonances are referred to, appropriately, as "peak fitting" algorithms. The AMARES (advanced method for accurate, robust, and efficient spectral fitting) software is an example of one system that performs peak fitting in this manner in the time domain (25).

Since many metabolites contain multiple protons, the MRS of a metabolite may consist of multiple Lorentzian-Gaussian resonances, referred to as "multiplets". Mathematically, a basis set can be established for each metabolite consisting of a combination of known multiplet parameters. For single-voxel MRS, the gold standard

software for spectral fitting is LCModel, a software package developed by Stephen Provencher in 2001 that utilizes basis sets to model spectra (26). By modelling all the resonances present in a spectrum, it is possible for LCModel to fully parameterize the spectrum. However, challenges arise when certain chemicals are present in very low concentrations or when resonance peaks overlap.

Other algorithms seek to only use spectral shape models for metabolites of interest and separately model the "background" or "baseline" signal of the spectrum. The quantitation based on semi-parametric quantum estimation (QUEST) algorithm operates on time-domain spectra using the Lorentizan-Gaussian lineshape for metabolites and applies singular value decomposition (SVD) to model the smooth components of the baseline signal (27). More recently, the Totally Automatic Robust Quantitation in NMR (TARQUIN) algorithm utilizes Hankel SVD to remove the water signal from spectra, thereby reducing much of the background signal, and a non-linear least squares method for separating out metabolite basis sets (28). For whole-brain sMRI obtained using the EPSI sequence, the FITT program of the Metabolite Imaging and Data Analysis System (MIDAS) operates on frequency-domain spectra and uses Lorentzian-Gaussian models for the peak components of a spectrum, and some sort of smooth function such as splines or wavelet reconstruction for the baseline (29,30). Liang et. al. expanded the idea of a basis set as a set of metabolite resonances to a set of tissue distributions containing multiple metabolite each (referred to in their work as "subspaces"); e.g. – a basis set could include a priori knowledge of the observed spectra for "normal neurons", "glioblastoma", and "lipid" (31).

In each of these algorithms, a common theme is the need for iterative processes such as expectation-maximization or the Levenberg-Marquardt algorithm to decompose the spectra into its constituent parts. These algorithms have a tradeoff between computational time and accuracy: an increased number of iterations are more likely to lead to convergence, assuming the models selected for spectral shape and background signal are appropriate. In general, these iterative approaches cannot be easily parallelized, and so while software such as LCModel are excellent for single and multi-voxel spectroscopy, they do not scale efficiently for whole-brain spectroscopy. MIDAS, which is a commonly used software for data acquired using EPSI, can take 40-60 minutes (even with powerful multi-threaded hardware) to perform spectral fitting on a single study, placing a practical constraint on implementation of this technology in the clinical workflow.

1.3.2 Inter-subject Comparison

MRS by its nature is not directly measuring metabolite concentrations but rather the radiofrequency (RF) signal of excited protons returning to their ground state. While this is correlated with concentration, the only way to get true concentrations is to calibrate to some reference signal with known concentrations. This can be done via an external phantom (and adjusted for partial volume within voxels and for the load on the excitation/receiver RF coils) (32,33), or by normalizing metabolite signals with the intracellular water signal (34). The latter technique, which is also implemented in MIDAS when using the EPSI sequence, relies on published reference values for the water content of voxels in T1-weighted MRI. Thus, absolute concentrations generated using this methodology do not have molar units, but rather are referred to (in MIDAS) as "institutional units"; this means that different studies obtained on the same instrumentation

are comparable, but not necessarily comparable for subjects scanned on different hardware (18). Metabolite ratios, by virtue of being unitless, are comparable across subjects regardless of instrumentation.

While normalization via intracellular water and metabolite ratios allow comparison of voxel values between subjects, there is still the issue of inter-subject variability that needs to be taken into count before one can identify voxels as being "within normal limits" or abnormal. Metabolite values are known to vary based on location within the brain (17,35), on subject age (36,37), in the presence of global inflammation (38), and even in healthy tissue when there is pathology elsewhere in the brain (39-41). A common method for internally normalizing a subject's metabolite map is to divide all voxels by the mean metabolite value in normal-appearing white matter (NAWM), a region defined radiologically as white matter tissue in the hemisphere of the brain without pathology. This technique is used to generate several advanced metrics for MRS, such as the choline-to-NAA index (42) and the Cho/NAA abnormality index (also called "relative Cho/NAA") (20). It is the Cho/NAA abnormality index that was found by Cordova et. al. to be the most correlated with histological presence of tumor cells, as described previously. The NAWM is typically defined by neuroradiologist segmentation, a process which is time intensive and requires manual intervention by a physician. Thus, to improve the clinical use of MRS/sMRI using internal normalization, improved algorithms that do not require manual segmentation are needed. At the time of writing, the MIDAS software suite has a beta version of automated NAWM segmentation based on clustering of tissue into grey matter, white matter, and cerebrospinal fluid.

1.3.3 Quality Analysis

In MRS and particularly in whole-brain spectroscopic imaging, the imaging field-of-view can consist of several thousand spectra. Quality analysis (QA), also known as artifact detection, describes the process of systematically identifying spectra that are artifactual – of poor quality – such that accurate quantitation cannot be determined from them and they should be excluded from any clinical decision making. In a 2004 review article, Kreis performed a systematic review of the types of artifacts that commonly affect MRS, including poor signal-to-noise, motion correction, lineshape distortions, partial volume, Eddy currents, insufficient water or lipid suppression, and errors in software processing (43). While this review was written primarily for single voxel and multivoxel MRS, whole-brain sMRI is perhaps even more sensitive to some of them. For example, during acquisition, shimming is performed to try and smooth out any magnetic field inhomogeneities which can lead to off-resonance signals; but with the large volume of and the heterogenous tissues within the field-of-view in whole-brain scans, inhomogeneities tend to play a larger role.

Manual inspection is still considered the gold standard for identifying poor quality spectra (43), though several algorithms for quality analysis have been developed to systematically identify poor quality spectra without expert review. This is critically important for whole-brain spectroscopy, where manual inspection of thousands of spectra is impractical. The first type of QA algorithm operates directly on the acquired data and assess statistical metrics such as magnitude of spectra or metrics of variability (e.g. kurtosis) (44). These metrics have the advantage that they are agnostic to the performance of spectral quantitation algorithms and can remove the most egregious artifacts such as

voxels obtained within areas of necrosis or outside the brain, but are not sensitive enough to identify other types of artifacts. The most common set of QA algorithms rely on metrics of spectral quantitation, such as the linewidth of peaks, the Cramer-Rao bounds of the fit, or the root-mean-squared error (RMSE) of the fit (30,45). These algorithms are more sensitive since they incorporate information from known spectral models and can therefore identify when the acquired signals deviate from the expected model; however, because they are dependent upon spectral quantitation, any errors in spectral processing can lead to erroneous identification of artifacts during QA. A third class of QA algorithm utilizes machine learning techniques to automatically classify spectra as being of "good" or "poor" quality; these techniques are described in the next sections.

1.4 Machine Learning

In traditional computational techniques, programmers will write code to perform specific operation on an input data set in order to produce some meaningful output. Machine learning is a broad term used to describe a variety of statistical and computational techniques that seek to learn what those specific operations should be; in other words, both the input and output are known, but the relationship between them needs to be learned. In general, there are two types of machine learning techniques that can be applied to tasks such as input classification – supervised and unsupervised techniques (46). Supervision refers to how a machine learning algorithm is trained, e.g. is it guided by known output (supervised), or can the output be inferred from some properties of the input (unsupervised). While machine learning encompasses a large field of study, in this work

we focus specifically on a type of system known as the neural network. Neural networks are most often trained using a supervised technique, and the rest of this section assumes this paradigm; in Chapter 4, additional discussion will be presented on how neural networks can be trained in an unsupervised fashion for a specific type of task.

Neural networks, also known as artificial neural networks (ANNs), are systems of linear operations designed to have plasticity, or the ability to update themselves via by analysing their inputs and outputs, so-called because of their similarity to how the brain works (in a simplified fashion) (47). ANNs are composed of nodes called neurons; each neuron receives input signals from one or more neurons, combines these inputs together, and if it reaches some activation threshold, passed an output signal to other neurons. By combining thousands of inter-connected neurons in this fashion, complicated non-linear functions can be approximated by the simple linear combination of signals. The ANN will undergo "training," wherein the network is presented with some input and the expected output. Mathematical operations such as gradient backpropagation can be used to tune each node – which other nodes it should receive input from, how much weight it should give to each input signal, and what its activation threshold should be – such that the network can turn the input into the expected output. Over time and many input-output pairs, the network will eventually converge into a steady-state wherein it can reproduce all input-output pairs in the training data set. With a sufficiently large training set and sufficiently large number of neurons, it is possible for an ANN to learn any possible input-output relationship. However, this comes at a large computational cost, with the number of possible inter-neuron connections increasing exponentially with the number of neurons in a network, and gradient backpropagation needing to be calculated for each connection. ANNs were an area

of active research for decades – a Google Scholar search reveals over 231,000 published articles from 1980 to 2010 containing "artificial neural network" – but these were limited by the computational resources available. Two main advances have driven more recent research in neural networks. The first is a drastic increase in computational power that occurred in the 2010s with the rise of commodity graphical processing units (GPUs), processors that were designed initially for computer graphics but contained hundreds of independent processing cores that could run backpropagation in parallel. The second is the development of the convolutional neural network (CNN), which rely on the fact that input data typically have some structure (e.g. images) and so the fully density of ANN connections is not necessary to capture this structure (48). Instead, kernels of much smaller size than the input can be convolved with input data in order to extract certain features from the data. In a CNN, input data are passed through several "layers" of convolution (and potentially some other layers such as traditional ANNs) to generate an output; during training of the CNN, the values of the kernels are updated. Thus, instead of needing an exponentially increasing number of inter-neuron connections as the number of neurons increases, in a CNN, the increase in complexity is approximately linear with the number of layers. With increasing computational resources such as GPUs and the ability to build more complex models, "deep learning" (referring to the increasing number of layers that data are passed through) has become a burgeoning area of academic research. Deep learning networks have shattered the performance benchmarks against both deterministic models and older machine learning techniques in many fields, from natural language processing to DNA mutation analysis to playing video games (49,50). While much of the initial work in machine learning was focused on everyday imaging, medical imaging can also benefit

greatly from these advances. Deep learning has also recently been shown to be a powerful method for extracting complicated feature sets from medical images (including MRI) in a fully automated fashion (51,52).

In both supervised and unsupervised training of neural networks, there is some information is provided to the network *a priori* and other information that must be inferred by the network. The terminology for this is often referred to as the "input feature vector" or "input features". Depending on the problem at hand, it may be beneficial to put in more or less data into the feature vector with the following qualification – the choice of features biases the network. For example, in a classification task based on medical images, the simplest approach would be to just provide the images as input and let the network try and extract all salient information. However, additional data can be input as additional features, such as statistical properties of the image or information about the subject whose image it is; the network no longer needs to infer these features and can instead use them directly in decision making. However, the network is now biased to rely on these features instead of determining computationally what are the best inherent features of the image. In practice, information should generally be provided if it is desirable to bias the network in a particular fashion. This is task of feature engineering – selecting what features should be input into the network based on knowledge of the task at hand. In Chapter 2, a CNN is developed for spectral artifact filtering that explicitly incorporates some information about the resonances of MRS but not any features that were generated by other algorithms (e.g. a specific spectral fitting algorithm). In this way, some knowledge about the fundamental properties of MRS can be used to bias the network, but the network is independent of other engineered features and can therefore be applied in the absence of those features (e.g. if a different

spectral fitting algorithm is used). Similarly, in Chapter 4, an unsupervised training architecture is developed that explicitly incorporates knowledge about the shape of spectral peaks, forcing the network to be confined to those spectral shapes.

1.5 Machine Learning for Magnetic Resonance Spectroscopy

In recent years, several MRS groups have adapted machine learning techniques for the imaging modality, focused primarily on acquisition (e.g. compressed sensing), spectral quantitation, and quality analysis. This section will cover the recent literature for the latter two foci as they are most relevant to this dissertation.

1.5.1 Quality Analysis

While the QA algorithms described previously can identify the majority of poor quality spectra (in our lab's experience, perhaps 70-80% of all artifactual spectra), several artifacts remain and cause issues with visualization and target generation. Since manual inspection by MRS experts is considered the gold standard, a new set of QA algorithms seek to generate tools that can mimic the expert's *gestaltism*. In 2008, Menze et. al. used random forest classifiers to separate high and low quality spectra per expert reviewers; of note, the feature vector used by the random forest was the magnitude of spectrum within a fixed resonance interval (44). That same year, Wright et. al. using independent component analysis to decompose spectra into several features, which were then fed into a support vector machine trained with expert classifications of spectral quality (53). Pedrosa de Barros et. Al. showed excellent performance for a spectral classifier based on a random

forest trained using manually generated features from both the time and frequency domain signals as input – peak SNRs, non-peak SNRs, and first- and second-order statistics (54,55). Finally, Kyathanahally et. al. demonstrated near-perfect removal of a single type of artifact – ghosting, the presence of spurious echos leading to off-resonance peaks – using convolutional neural networks (56,57). Collectively, these projects have shown that machine learning can indeed be useful in artifact detection for MRS, but also that there is much work to be done. Specifically, machine learning has not been applied to whole-brain data collected using the EPSI sequence.

1.5.2 Spectral Fitting

As described in Section 1.3.1, spectral fitting can be a computationally intensive task and is prone to errors when signal quality is poor or in the presence of artifacts. While there is a plethora of research in the use of various optimization techniques to improve spectral fitting, there has been limited work to date on the application of machine learning techniques to this task (see Table 0.1 for a summary).

In general, there have been two tasks that this body of work has tried to solve: determination of metabolite concentrations (relative concentrations), or actual fittings that include a model of the baseline and each metabolite. Hiltunen et. al. used an ANN in 2002 and trained it to automatically determine the metabolite concentrations of single in vivo spectra from patients with gliomas, and observed ANN performance to be comparable to that produced by lineshape fitting models (58). In 2006, Bhat et. al. also used an ANN architecture, but to actually model peaks using Lorentzian-Gaussian functions rather than predicting metabolite areas (59). Thus, their model was firmly contextualized within the

constraints of resonance physics. However, it was limited in that it required the baseline signal to be removed by other means, and for the specific bounds of the peak signal to be identified. More recently, Das et. al. created two separate architectures – one based on a random forest and the other using a multilayer perceptron (MLP, a type of multilayer ANN) – predict metabolite concentrations on 2D MRSI data sets (60,61). They observed performance comparable with the gold-standard LCModel software, but in a fraction of the time. To date, however, no machine learning model has been published that can provide full spectral fittings (e.g. constraining to spectral lineshape models) that include a model of the baseline signal and any phase shifting that may be present. Such a model will be necessary if it were to replace existing software such as LCModel, jMRUI, or MIDAS.

Table 0.1: Summary of literature on machine learning for spectral fitting

Author	Data	Technique	Training	Output
Hiltunen, 2002 (58)	In vivo spectra	Artificial neural network	Lineshape fitting	Metabolite concentrations
Bhat, 2006 (59)	Baseline-corrected 2D PRESS MRSI	Radial basis function neural network	Unsupervised	Fitted spectra
Das, 2017 (60)	Simulated and in vivo 2D spectra	Random forest	LCModel	Metabolite concentrations
Das, 2018 (61)	Simulated and in vivo 2D MRSI	Multi-layer perceptron	LCModel	Metabolite concentrations

1.6 Specific Aims

While sMRI has been well established in the research arena, the clinical adoption of the technology has been limited due to several technical challenges. While the EPSI sequence has reduced acquisition times to a reasonable 15 minutes, post-processing of the data from the scanner into metabolite heatmaps showing in Figure 0.2 and Figure 0.3 requires several computationally-intensive steps. The presence of spectral artifacts – erroneously calculated metabolite concentrations due to magnetic field inhomogeneities – remain a concern for accurate volumetry, which is especially important if sMRI is to be used for treatment planning (43). Due to the complexity of its data, sMRI volumes cannot be visualized or analyzed using standard clinical imaging tools, and thus the incorporation of this information into the clinical workflow is limited. Finally, there are known variations in basal metabolism between patients, and this inter-patient variability needs to be addressed when determining practice guidelines for sMRI.

To address challenges, we have developed a software framework that incorporates the latest knowledge of machine learning and high-performance computing to bridge the gap in the clinical utilization of sMRI for radiation therapy guidance.

Aim 1: Develop preprocessing algorithms for spectral quality filtering using deep learning. Using deep learning, we will develop a spectral quality filter to remove aberrant spectra from sMRI volumes. We will develop an easy-to-use, web-based platform to allow MR spectroscopy experts to crowd-source "ground truth" data of spectral quality – a highly subjective metric – and use these data to train a neural network which can be applied to whole-brain sMRI volumes.

Aim 2: Develop normalization methods for personalized spectroscopic MR imaging of glioblastoma. To account for basal variations in metabolic levels, we will develop an automated and personalized normalization scheme for sMRI of patients with glioblastoma. The normalization scheme will be implemented based on an Gaussian mixture model optimized using expectation-maximization.

Aim 3: Develop a software framework capable of incorporating sMRI into the clinical workflow. To enable the use of sMRI in the clinical mainstream, we will develop a web-based application that will allow MR technicians to import raw spectroscopic data from an MR scanner and automatically produce clinically relevant 3D metabolite (and ratio) maps, and radiation therapy (RT) or surgical target volumes after co-registration with clinical anatomical MRI. This will serve as a platform for integrating algorithms in our research, such as the quality filter and normalization algorithm developed in Aims 1-2, and for coordinating multi-center clinical trials.

Aim 4: Improve efficiency of spectral fitting using a parametric deep learning model. Spectral fitting, the process of parameterized curve fitting on a raw spectrum, enables quantification of relative metabolite concentrations. Current algorithms are based on iterative optimization of the spectral baseline and peak models and can require upwards of an hour per 3D whole-brain scan, limiting the utility of sMRI in the immediate clinical setting. We will develop a single-pass neural network capable of operating in parallel on thousands of spectra to perform rapid fitting of spectra. We will develop an unsupervised learning architecture that incorporates models of spectral lineshape and baseline such that the resulting spectral fitting is fully constrained within known physics.

A CNN FOR SPECTRAL ARTIFACT FILTERING

1.7 Author's Contribution and Acknowledgement of Reproduction

This chapter is reproduced from SS Gurbani, E Schreibmann, AA Maudsley, JS Cordova, BJ Soher, H Poptani, G Verma, PB Barker, H Shim, and LAD Cooper. *A convolutional neural network to filter artifacts in spectroscopic MRI*. Magn Res Med, 2018. 80: 1765-1775. Permission has been obtained from the publisher, John Wiley and Sons. License number 4510301418809.

SSG designed the data collection tool. AAM, JSC, BJS, HP, GV, PBB, and HS provided expert ratings to train the neural network. SSG and LADC designed the neural network architecture. SSG, ES, HS, and LADC performed the analysis of the data. SSG wrote the manuscript with feedback from all authors.

1.8 Abstract

1.8.1 Purpose

Proton spectroscopic magnetic resonance imaging (sMRI) is a non-invasive modality capable of generating volumetric maps of in vivo tissue metabolism without the need for ionizing radiation or injected contrast agent. sMRI has been shown to be a viable imaging modality for studying several neuropathologies. However, a key hurdle in the clinical adoption of sMRI is the presence of spectral artifacts that can arise from a number of sources, possibly leading to false identification of lesions.

1.8.2 Methods

A deep learning model was developed that was capable of identifying and filtering out poor quality spectra. The core of the model used a tiled convolutional neural network (CNN) that analyzed frequency-domain spectra to detect artifacts.

1.8.3 Results

When compared to a panel of spectroscopy experts, our CNN achieved high sensitivity and specificity with an area under the curve of 0.95. A visualization scheme was implemented to better understand how the CNN made its judgement on individual spectra, and the CNN was embedded into a pipeline capable of analyzing whole-brain sMRI volumes in real-time.

1.8.4 Conclusion

The fully-automated method for assessment of spectral quality provides a valuable tool to support clinical sMRI studies for use in fields such as adaptive radiation therapy planning.

1.9 Introduction

Proton magnetic resonance spectroscopy imaging (sMRI) is a quantitative imaging modality that measures endogenous metabolite concentrations *in vivo* (22,62). Recent advances in sMRI protocols, such as the development of the 3D echo-planar spectroscopic imaging (EPSI) sequence and improved post-processing methods have enabled whole-brain acquisition with spatial resolutions on the order of one centimeter, thereby improving the utility for diagnostic and treatment planning applications (17,18,20,63). Three metabolites of key importance in the evaluation of patients with glioblastoma, specifically choline (Cho), N-acetylaspartate (NAA), and the Cho/NAA ratio. The Cho/NAA ratio is widely used for depiction of tumor volumes and infiltration due to increased contrast as a result of the opposite changes of these metabolites in the tumor (20). A key challenge to the analysis and interpretation of sMRI volumes is the presence of artifacts due to poor spectral quality (64). Artifacts arise from magnetic field inhomogeneities, subject movement, or improper water and lipid suppression, yielding reduced peak signal-to-noise ratios and distorted and broadened line shapes that lead to difficulties in quantification of the metabolite peaks (30,64). Visually, artifacts may appear as foci of hyper- or hypo-intense signal (Figure 0.1), which can lead to false interpretation. For treatment planning purposes, it is especially important to obtain accurate volumetry of target pathology. Currently, confirmation of true metabolic abnormality requires manual review of spectra by experts. However, with several thousand spectra in a single whole-brain volume, manual review is impractical. To adopt whole-brain sMRI into the clinical workflow, it is therefore necessary to develop automated methods for assessment of spectral quality.

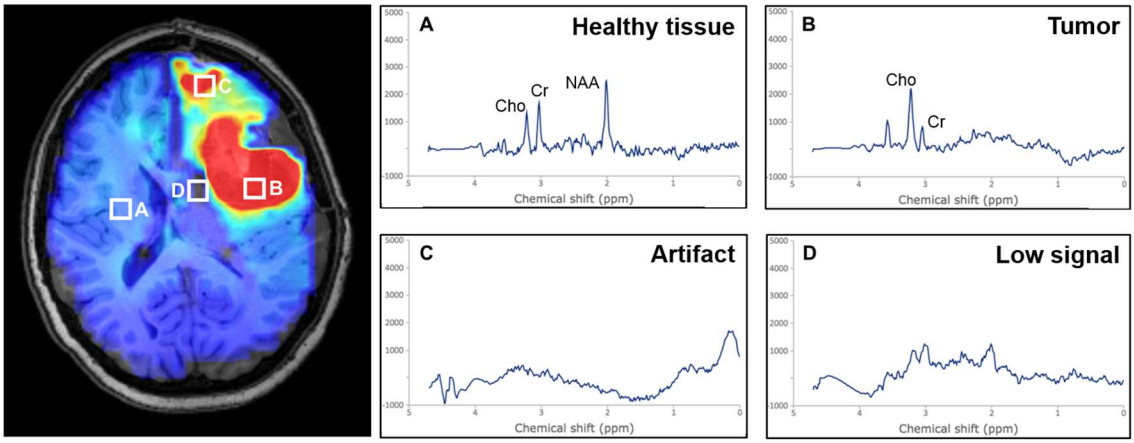


Figure 0.1: Sample gallery of sMRI spectra. Artifacts in spectroscopic MRI (sMRI) arise due to several causes, and can lead to false interpretation of pathology. (A) Healthy tissue shows a relatively low choline to N-acetylaspartate ratio (Cho/NAA), whereas (B) tumor shows an elevated ratio, appearing as hyperintense on a Cho/NAA map. Artifacts can arise in tissue boundaries and in areas with poor lipid or water suppression, and can result in either (C) hyperintense lesions or (D) dropout of signal. Accurate volumetry of pathology is critical if sMRI is to be used for treatment planning.

Several approaches have been developed to filter poor quality spectra from sMRI datasets, including exclusion criteria based on peak linewidths (17), reliability testing (65), Cramer-Rao bounds (45), and machine-learning techniques such as random forest classifiers (44,54). In each of these algorithms, the classification of spectral quality is based on a collection of "engineered" features as input; these features are categorical or ordinal vectors that seek to summarize the information encoded in each spectrum. Some features are derived directly from the raw data, such as the magnitude of each point in the spectrum or statistical metrics of variability (e.g. kurtosis, skewness) (44), or derived from analysis of spectral fitting, such as singlet linewidths and Cramer-Rao bounds. Common to all these approaches is that the definition of features is required prior to performing any analysis.

Thus, these features capture criteria that MR experts explicitly believe are important to spectral quality, and machine learning algorithms built using these engineered features have shown promise as spectral quality filters (44,54,66). In contrast to systems that model expert beliefs explicitly, deep learning broadly defines a category of machine learning algorithms that can learn underlying features from raw data without the need for any *a priori* definition of such features, and have been able to shatter benchmarks in natural language processing, medical image segmentation, survival analysis, and identification of pathology in medical images (67-71). Deep convolutional neural networks (CNNs) in particular are well suited to analyzing waveforms similar to raw spectra (72,73), and have only recently been applied in the context of MR spectroscopy (56). CNNs consist of sequential layers of convolution, downsampling, and logistic regression that ultimately learn the underlying features of the data that lead to a desired endpoint, in this case whether or not a spectrum is suitable for analysis. Training a CNN involves updating the coefficients of the convolutional layers to reach the endpoint. However, designing the architecture of a CNN – e.g. the number of layers and the size of the convolution kernels – often relies on trial and error. A new framework to determine optimal architecture parameters is known as Bayesian optimization, which performs non-linear optimization without the need for defined derivatives (74).

In this work, a CNN was developed to learn whether or not a spectrum has sufficient quality to be used for clinical assessment. A state-of-the-art Bayesian optimization technique was used to automatically tune the network architecture and train the network on data from patients with glioblastoma. A framework is provided that enables users to visualize the rationale behind the CNN's classification of individual spectra. Finally, a

software pipeline was implemented that enables real-time filtering of whole-brain sMRI volumes.

1.10 Methods

1.10.1 Image Acquisition and Processing

sMRI data were available from a database of patients with glioblastoma previously enrolled in a phase II clinical trial (75) who received post-surgical scans. All scans were conducted in a 3T MRI scanner with a 32-channel head coil (Siemens Medical) and were obtained following surgical resection but prior to the start of radiation therapy and chemotherapy. Anatomic volumes obtained used a T1-weighted (T1w) magnetization-prepared rapid gradient echo pulse sequence (TR = 1900 ms, TE = 3.52 ms, 256 x 256 matrix, flip angle (FA) = 9°). A whole-brain 3D echo-planar spectroscopic imaging sequence (EPSI, TR = 1551 ms, TE = 17.6 ms, FA = 71°, final matrix size of 64 x 64 x 32) yielding a final voxel volume of 1.5 cm^3, was obtained during the same scanning session. Both sequences were obtained at a +15 degree tilt in the sagittal plane from the anterior commissure-posterior commissure line in order to capture the entire cerebrum while minimizing acquisition in the clivus, sinuses, and retro-orbital fat. An oblique saturation band was placed in the sagittal plane from the optic chiasm to the cerebellum to suppress signal from those regions. Image reconstruction and formation of metabolite images were carried out using the Metabolite Imaging and Data Analysis System (MIDAS) package (17,18). Briefly, this processing includes spatial reconstruction, frequency alignment, B0 field correction, co-registration of the T1w and EPSI volumes, registration

of the T1w volume to an anatomic atlas, lipid suppression, spectral fitting, and normalization with internal water signal to produce relative concentrations of metabolites. Additionally, pre-filtering of data using algorithms built into MIDAS were applied to all data to replicate the workflow currently used in clinical studies. First, a mask based on an anatomic atlas was applied to exclude voxels outside of the brain, which drastically reduces the number of voxels to be analyzed. Voxels with a water linewidth > 18 Hz, as calculated from T2 decay, were removed prior to spectral fitting to save computation time. After fitting, voxels with a metabolite linewidth > 18 Hz were also removed; this step served as an initial filter to remove spectra known to be of poor quality prior to visual review. A representative volume contained 10,298 voxels after filtering, however, since the data is also interpolated and smoothed in each dimension during reconstruction it is estimated that approximately 1,280 independent spectra remain within the brain volume. Spectra were then randomly sampled from a grid with a skip factor of 2, including both regions of tumor and healthy tissue, and exported for analysis. A total of 8,894 spectra collected from 9 patients with glioblastoma were collected.

1.10.2 Data Labelling and Consensus

A custom web platform for multi-rater annotation of spectral quality was developed and used to label the exported spectra for training and testing the artifact detection network. Online Spectroscopic Classification and Review (OSCAR) was written in PHP – a language for server-side scripting of webpages – and enables multiple MR spectroscopy experts to visualize and review spectra, as seen in Figure 0.2. Each expert is presented with a random subset of spectra and shown the chemical-shift spectrum along with a cross-haired T1-weighted MR image depicting the spectrum's anatomic location. The user

classifies the spectrum's quality as "Good", "Acceptable", or "Poor", based on their expert opinion.

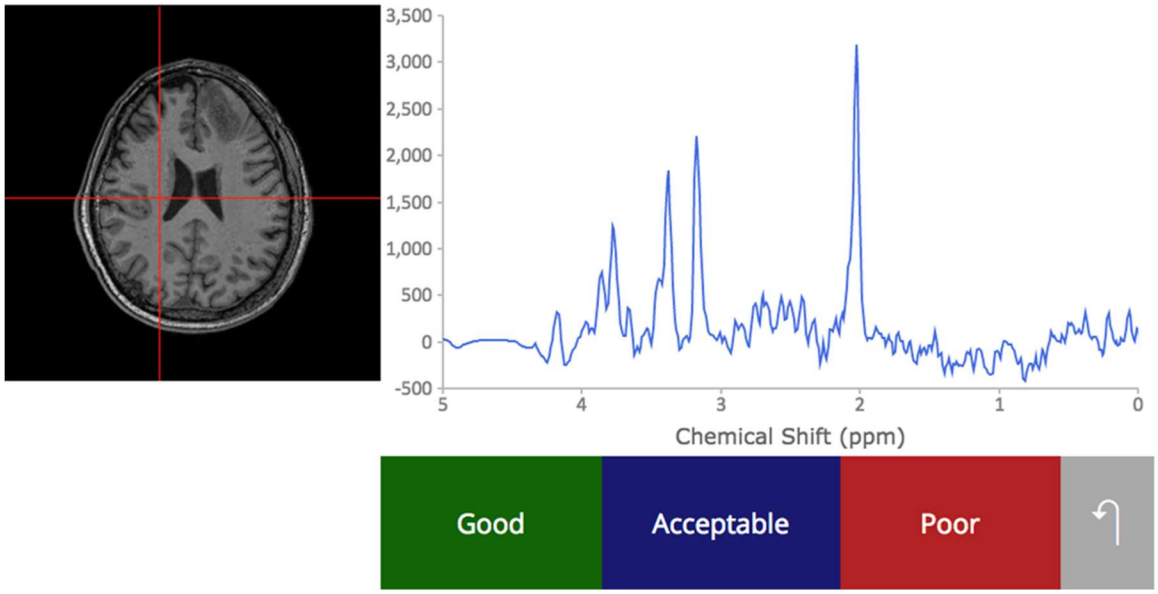


Figure 0.2: A web interface for MRS expert ratings. To collect ground truth data for machine learning classifiers based on spectral quality, we developed a web-based interface for MR experts to use.

Each spectrum was presented to three randomly chosen users from a pool of experts (authors AAM, JSC, BJS, HP, GV, PBB, and HS) so that an expert consensus could be measured; a total of 8,894 spectra were classified in triplicate. Consensus was based on a majority vote decision rule; if all three voted differently or if there was a large discrepancy in votes (e.g. 2 "Good" and 1 "Poor" or vice versa), the spectrum was discarded. Table 0.1

includes the results of labeling and consensus; a total of 427 spectra (4.8%) were discarded due to large discrepancy amongst experts. The consensus labels of “Good” and “Acceptable” were merged together into a singular “Good” class, yielding a class proportion of 72% “Good” and 28% “Poor”. The final set of 8,467 spectra were then randomly split 80:10:10 for training, validation, and testing subsets, with class proportions maintained.

Table 0.1: Consensus rating of spectral quality by a panel of MRS experts.

Reviewer Labels			Consensus Label	# of Spectra	% of Total
G	G	G	Good	360	4.05
G	G	A	Good	914	10.28
G	A	A	Acceptable	1597	17.96
A	A	A	Acceptable	1672	18.80
A	A	P	Acceptable	1448	16.28
A	P	P	Poor	1120	12.59
P	P	P	Poor	1356	15.25
G	G	P	*Discarded*	23	0.26
G	A	P	*Discarded*	355	3.99
G	P	P	*Discarded*	49	0.55
			Total Spectra:	8894	
			Spectra discarded:	427	4.80
			Final data set:	**8467**	95.20
			Training set:	6767	
			Validation set:	850	
			Testing set:	850	

1.10.3 Network Architecture

A CNN for spectral quality was developed using the TensorFlow (Google, Inc.) framework on a Windows workstation with two Titan X graphical processing units (GPUs). GPUs are highly optimized for parallel computations and enable rapid development of neural networks. A high-level overview schematic of the CNN is displayed in Figure 0.3a. It takes the 512-point real component of a spectrum as input. First, the spectrum is normalized to values between 0.0 and 1.0; for this step, a histogram of amplitudes of all spectra in the training data set was computed, and the 1st and 99th percentile values were used as the normalizing bounds. Using these bounds, all spectra are normalized to the same scale. The normalized spectrum is then split into six regions, called "tiles". Three of these tiles were based on the known location of resonance peaks for the three largest metabolite peaks present: Cho (3.2 ppm), Cr (3.0 ppm), and NAA (2.0 ppm). Given the short TE that the data were acquired at, these three tiles are primarily composted of a Lorentzian-Gaussian singlet as opposed to the slowly-changing spectral baseline and smaller overlaid macromolecule singlets found in other tiles (30). Since these tiles have different shapes, the tiled architecture effectively enables six parallel networks to be trained synchronously. Each tile is passed through a series of convolutional and max-pooling layers. Layers 1 and 2 each perform 32 convolutions on the data, and layers 3-7 each perform 64 convolutions. Since CNNs operate best on low-amplitude data, outputs from each convolution are passed through a parametric rectified linear unit (PReLU), which restricts output range (76). Max-pooling downsamples the output of each convolution by a factor of two, reducing the size of the data going into the next layer. After these 7 layers, outputs from each of the six tiles are concatenated together and passed through two fully-connected layers of 128 nodes.

Finally, logistic regression and softmax (77) operations are performed to yield a single scalar output within the range [0.0, 1.0]. The CNN is trained so that this output is the probability of the input spectrum being classified as “Good”.

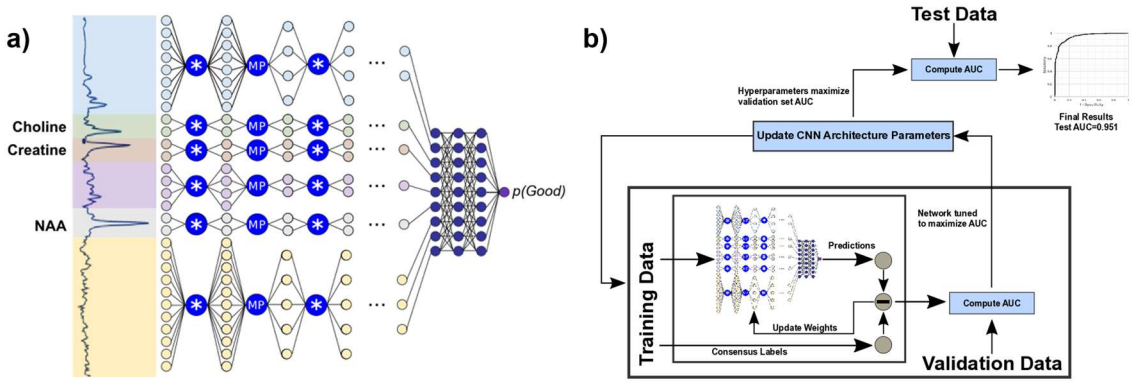


Figure 0.3: Schema of the spectral quality CNN and its training process. (a) A high-level overview of the convolutional neural network for spectral quality analysis. Input spectra are split into six tiles and passed through a series of convolution (*) and max-pooling (MP) layers, then concatenated and passed through fully connected layers to generate a scalar output of spectral quality. (b) Bayesian optimization is used to iteratively optimize architecture hyperparameters.

Training the CNN involves optimizing the coefficients of all the convolution kernels and fully-connected nodes so that the output matches the consensus of the training data generated in OSCAR. If the consensus output was “Good”, the spectrum is given a class label of 1, and if the consensus output was “Poor”, the spectrum is given a class label of 0. The cost function to be minimized by a first-order optimizer is defined as the cross-

entropy error of the class label and the output probability of the CNN (78). Spectra are passed in batches of 250 spectra through the network and the cost function is computed; gradient backpropagation then adjusts the network coefficients. For each batch, spectral from the training data set are randomly sampled without replacement; an epoch is defined as one full pass-through of the training set. After each epoch, a receiver-operator characteristic (ROC) curve is computed on the validation (n=850) sets, and the area under the curve (AUC) is reported. Training continued through many epochs until the AUC of the validation set converges.

1.10.4 Bayesian Optimization

CNN architecture hyperparameters, e.g. the number of layers in the CNN and the size of the convolution kernels, can have a large impact on performance. These hyperparameters cannot be defined by linear mathematical functions and therefore do not have explicit derivatives to be used for gradient backpropagation. Because of this, they are outside the scope of neural network training and are often defined by manual trial and error, which can be both time-consuming and highly subjective. Bayesian optimization is a statistical technique that models the performance of an algorithm using Gaussian processes and iteratively seeks to optimize performance over the space of possible algorithm designs (74,79,80). CNNs can be treated as a complex mathematical function with some unknown underlying statistical model. Bayesian optimization assumes a series of Gaussian distributions can reproduce the underlying model and uses those to indirectly optimize the CNN; this technique has previously been successfully applied to tune architecture hyperparameters (71,74). A Bayesian optimization approach using the Spearmint framework is used in this work to tune the size of the convolution kernels and the dropout

fraction of the fully connected layers (74). The variable to be minimized for Bayesian optimization is defined as the converged AUC of the validation data set as described in the previous section (Figure 0.3b).

1.10.5 Gradient-weighted Class Activation Mapping

While deep learning algorithms such as CNNs extract predictive features from input data, these features do not have any semantic meaning, and it is not readily apparent how to interpret them in a manner similar to feature ranking by standard statistical techniques. A recent method known as gradient-weighted class activation mapping (GradCAM) attempts to visualize which components of a specific input contribute to that input's classification. GradCAM performs a gradient backpropagation from the output score to the final layer of convolution, which contains the highest level features detected by the network (81). These can then be visualized as heat maps over the original input, highlighting which components contributed most to the CNN's decision. While this cannot be generalized to all input, it enables insight into what the network is doing. We implemented GradCAM for spectral data in TensorFlow.

1.11 Results

1.11.1 Training and Validation

The CNN was trained over multiple epochs (complete passes through the training set) until the validation AUC achieved its peak value, an indication that additional training would result in overfitting. The Bayesian optimization framework was programmed to

maximize the peak validation AUC, and this along with a 2x2 contingency table (82) was reported for each parameter set. The final tuned parameters reported by Spearmint were the CNN learning rate and convolution kernel size. After Bayesian optimization was complete, the unused test set (n=850 spectra) was run through inference and probabilities were compared with reviewer labels to generate the final ROC curve shown in Figure 0.4a, with an AUC of 0.95.

To compare the classifications of the CNN against each human expert's classifications, an inter-rater agreement analysis was performed. A dissimilarity matrix, representing the percent disagreement between each pair of observers was calculated (Figure 0.4b). Multidimensional scaling a technique that transforms these pairwise distances into a two-dimensional map to further visualize the relative agreement between pairs of raters, was performed (Figure 0.4c) for the seven human raters (H1-H7) and the CNN on the full data set. In a multidimensional scaling plot, points that are closer together represent higher agreement between the two raters' decision on spectral quality. The geometric center of the human raters is also displayed on the plot. After training and tuning of the CNN were complete, it was run on the 427 spectra which the experts disagreed on, using a threshold of 0.7 on the CNN's output probability for classification. A summary of the results is presented in Table 0.2; of note, the majority of these spectra were classified by the CNN as having good quality.

Table 0.2: CNN classifications of the discordant data set.

Reviewer Labels	# of Spectra	CNN Classification		CNN Probabilities Mean (std. dev.)
		Good	Artifact	
G G P	23	23	0	0.998 (0.005)
G A P	355	308	47	0.871 (0.261)
G P P	49	31	18	0.686 (0.380)

Table 0.3: Feature vector generated by MIDAS used for training a random forest.

Spatial Location	**Choline Peak**	**Creatine Peak**	**NAA Peak**
X coordinate	Area	Area	Area
Y coordinate	Frequency	Frequency	Frequency
Z coordinate	Normalized concentration	Normalized concentration	Normalized concentration
Fit Parameters	**Quality Metrics**	**Other Metrics**	
Zero-order phase	Residual error	B0 Field Map	
First-order phase	Chi-squared statistic	Bias Field Map	
Metabolite Linewidth	MIDAS QMap*		

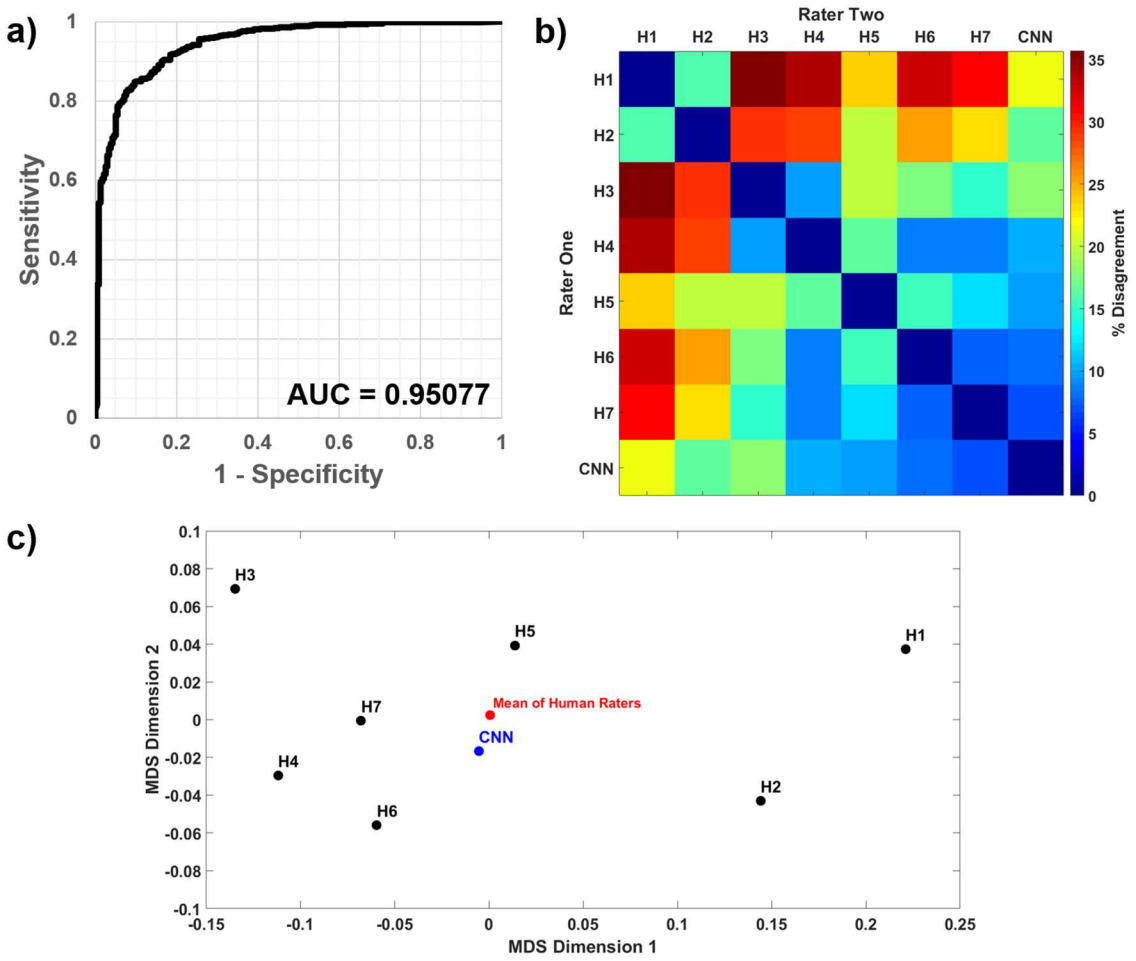


Figure 0.4: Spectra quality CNN performance versus MRS experts. (a) An unused test data set (n=850 spectra), with class proportions matching that of the full data set, was run through the CNN; comparing the output probabilities to ground truth resulted in a receiver-operator characteristic curve with an AUC of 0.951. The (b) dissimilarity heatmap and (c) a multidimensional scaling plot comparing sets of pairwise inter-rater agreement show that the CNN's similarity with any given human rater is within the ranges of inter-human rater similarity.

To evaluate the relative performance of the CNN against traditional machine learning with engineered features, a random forest was generated using the same input and ground truth data using the Statistics and Machine Learning Toolbox in MATLAB R2016b (The MathWorks, Inc.). Twenty features generated by MIDAS during the fitting process were used (Table 0.3), and the model achieved an AUC of 0.948 after optimizing random forest parameters. The ranking of features is shown in Figure 0.5.

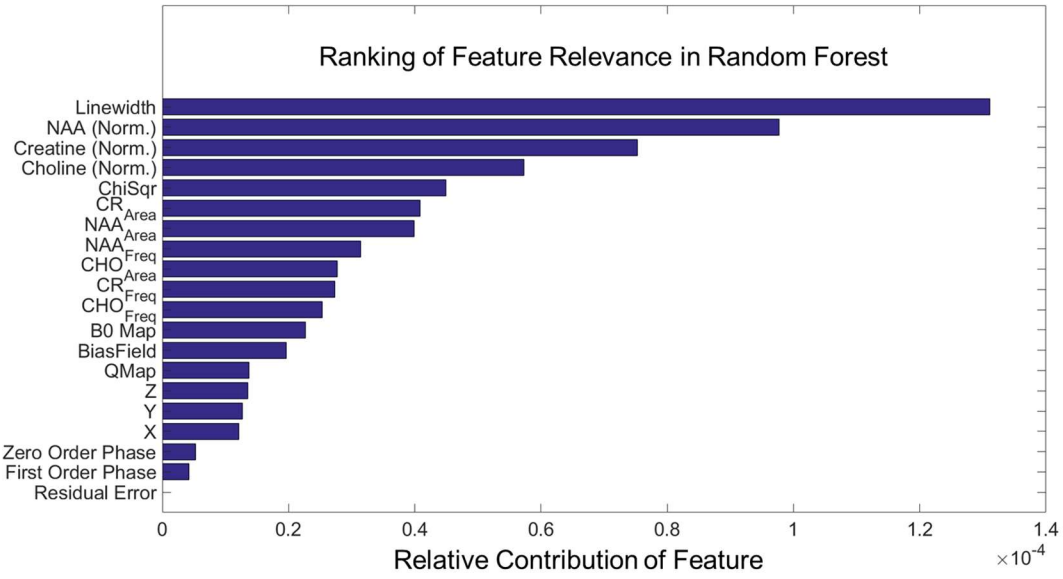


Figure 0.5: Ranking of random forest feature vector by relevance. The MIDAS features used in the random forest classifier are ranked by their relative strengths in determining the forest's classification decision. The random forest achieved an area-under-the-curve (AUC) of 0.94 when comparing with the consensus decision of an MR spectroscopy expert panel.

1.11.2 Gradient-weighted Class Activation Mapping

Representative spectra are shown in Figure 0.6 with overlaid heat maps generated by GradCAM, revealing regions of the spectrum that contributed most to the CNN's classification decision. Four representative spectra are shown, each with different causes leading to poor quality per the opinion of two of the expert spectroscopists. GradCAM highlights the regions of spectrum that correspond to those causes. To assess how the CNN makes a classification decision on good quality spectrum, an idealized "Good" spectrum is generated by taking the average of the amplitude all spectra classified as "Good" by the five readers (Figure 0.7), which increases the signal-to-noise ratio and reduces baseline variations. When GradCAM is run on this idealized spectrum, the heat map reveals that it is the regions outside of the metabolite resonance frequencies, e.g. the region of lipid and water signal, that contribute most to the CNN's classification.

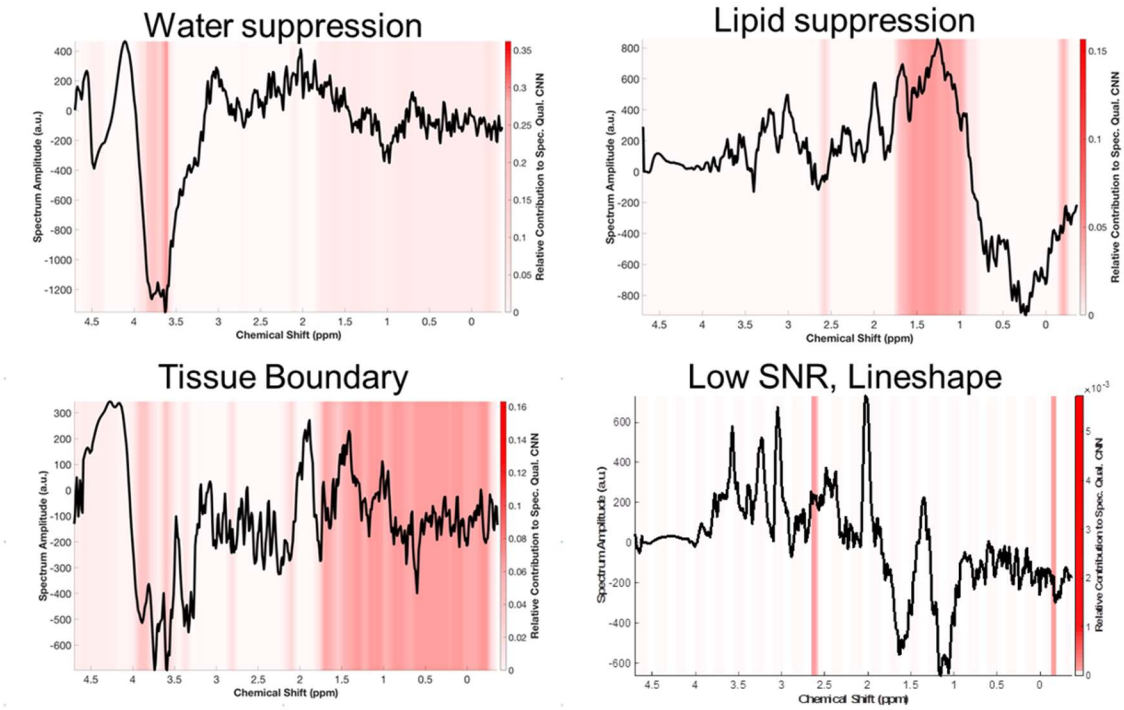


Figure 0.6: GradCAM analysis of representative poor quality spectra. Four spectra representative of the various phenomena that lead to artifacts were analyzed using GradCAM, a technique which produces a heat map of which portions of a specific input spectrum contributed most to the CNN's final decision. The results show that the CNN is focusing on appropriate regions for each of these scenarios. Of note, when there is low signal-to-noise ratio due to partial volume effects at a tissue boundary (bottom left), nearly the entire spectrum is detected.

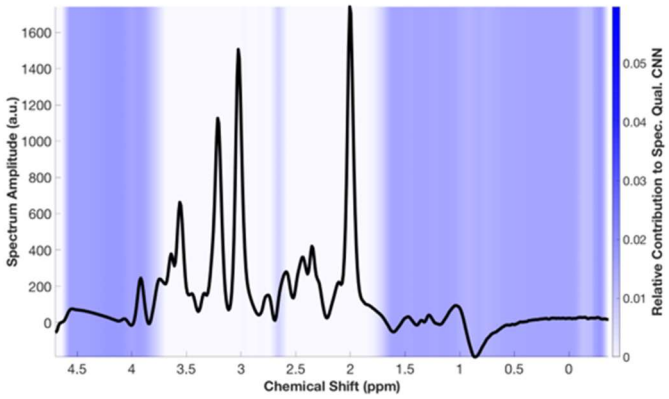


Figure 0.7: GradCAM analysis of an idealized good quality spectrum. An idealized "Good" spectrum, created by averaging all spectra classified as "Good," shows that the CNN focuses most on the regions outside of the metabolic peaks for decision making.

1.11.3 Whole-brain Pipeline

A pipeline was developed for filtering spectral artifacts on whole-brain volumes. Briefly, all spectra were exported to a binary format able to be read by TensorFlow and tagged with their voxel locations for reconstruction. The trained CNN was then loaded and inference is performed on all voxels in a whole-brain volume; a probability of "Good" quality is exported and reconstructed into a 3D volume representing probability. A threshold on this probability can then be selected and applied to all metabolite voxels, eliminating those below it from the map; a threshold of 0.70 was selected for the whole-brain images depicted here. Figure 0.8 shows a sample pre-radiation therapy Cho/NAA volume from a patient with a newly diagnosed left frontal glioblastoma following surgical resection of the tumor. Voxels with broad linewidths were already removed during data

processing in MIDAS with an initial filter based on water and metabolite linewidths. Cho/NAA elevation is observed posterior and medial to the surgical cavity, suggesting residual tumor. Additionally, in the unfiltered Cho/NAA maps, there appears to be residual tumor in the anterior tip of the left frontal lobe. However, inspection of the spectra in these voxels reveals that the spectra are of insufficient quality to make an accurate assessment of pathology. Examples of spectra from eliminated voxels are shown, taken from regions of cellular necrosis, low signal-to-noise-ratio, and insufficient lipid suppression. This whole-brain pipeline is implemented using the Python version of TensorFlow and is portable across multiple computer operating systems and hardware, including low and mid-tier GPUs. On a system with a low-end GPU (Nvidia GTX 1050 Ti), processing takes less than 2 minutes. Removing the GPU and running the pipeline only on a multicore CPU (Intel i7-6900K, 8 cores) also takes 2 minutes to process a whole-brain volume.

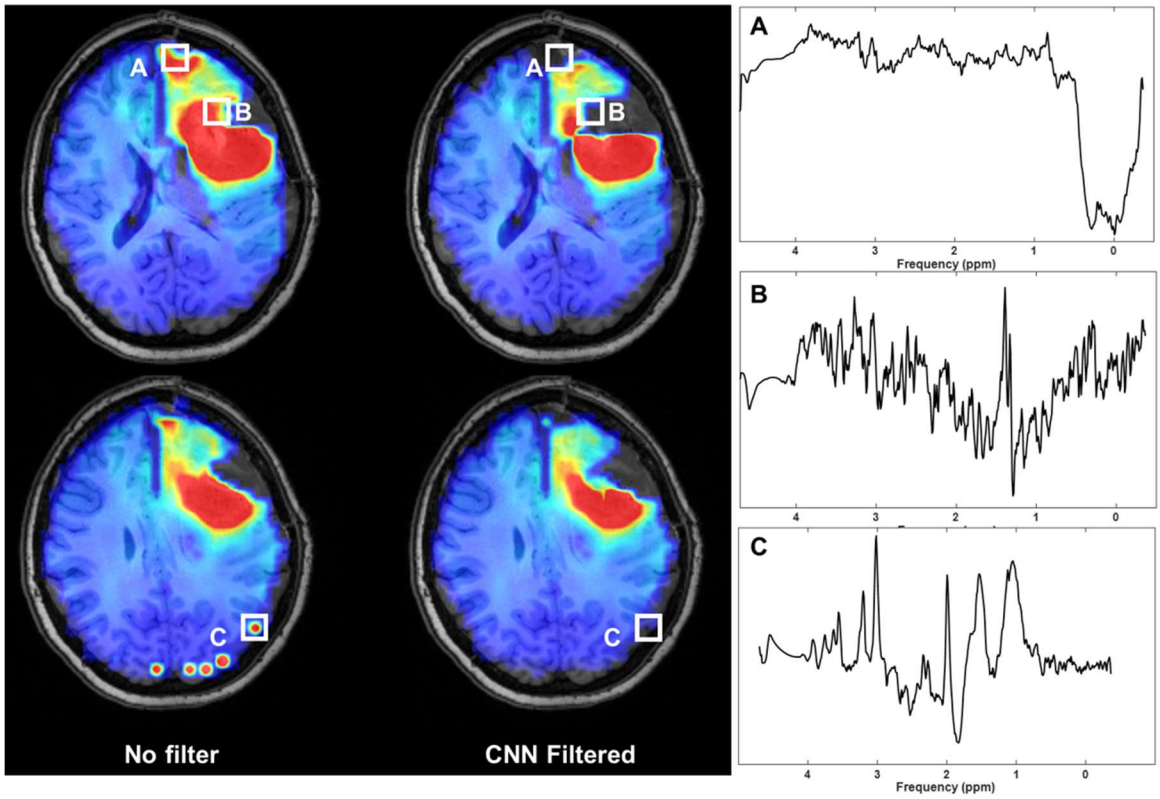


Figure 0.8: Application of spectral quality CNN to whole-brain metabolite heatmaps. The spectral quality CNN can be applied to whole-brain volumes in real-time to assist clinicians in making accurate decisions based on sMRI. In a pre-treatment assessment, the CNN filters out voxels in the necrotic tumor core, anterior frontal lobe, and multiple voxels in the occipital lobe which have poor quality.

1.12 Discussion

The CNN-based spectral quality filter developed in this study was found to perform well for detection of inadequate quality spectra, as labeled by a consensus decision of MR spectroscopy experts, with an AUC of 0.951, indicating the network has been tuned for both high sensitivity and specificity. Furthermore, the network is wrapped into a framework that enables rapid deployment of the filter into the clinical research workflow and can be applied in under 2 minutes to high resolution EPSI data with whole brain coverage. The accuracy achieved is similar to that reported in previous studies using machine learning for MR spectral quality analysis, such as those utilizing random forests with engineered spectral features (54,55,66). It is difficult to compare results across studies, due to variation in study design: how data was collected, the biases of the raters generating ground truth, and which parameters were chosen as features. To enable direct comparison of performance between the CNN and a machine learning algorithm with engineered features, a random forest was built on the same training and testing data collected in OSCAR. While the CNN demonstrates slightly better performance, a key advantage of the deep learning approach over the random forest is that does not require any engineered features generated by spectral fitting. As a result, it is independent of specific features derived from fitting algorithms and is compatible with a broader set of pipelines and workflows. Of note, the spectra in this work do undergo preliminary linewidth filtering in MIDAS prior to being processed by the CNN, which reduced the spectra needed to undergo spectral fitting and excluded spectra with broadened metabolite linewidths known to be of poor quality. This decision was made to use the *status quo* of spectral quality filtering as a starting point and to develop an algorithm that makes additional improvements to it. In

future workflows, applying filtering prior to spectral fitting could reduce the computation time for spectral fitting, a time-intensive process, especially since poor quality spectra may require considerably more processing during fitting, due to their abnormal shapes.

While artifacts can occur anywhere in the tissue parenchyma, they are more common at air/tissue boundaries due to local magnetic field inhomogeneities and near the periphery due to partial volume effects with strong lipid signals in scalp. It is well known that magnetic field inhomogeneity occurs in the inferior frontal lobe, anterior temporal lobe, and superior to the mastoid air spaces due to magnetic susceptibility differences between air and tissue. In patients who have undergone surgery, additional artifacts in EPSI data can occur due to magnetic susceptibility from craniotomy staples and possibly from hemorrhage. In this first iteration of the spectral quality CNN, this information is not taken into account insofar as it is not encoded into the spectrum itself. Future work could include the registration of sMRI volumes onto a common anatomical atlas and input the anatomical location of each spectrum into the CNN, either in terms of absolute location or relative distance from the nearest tissue boundary (i.e. ventricle, dura).

Additionally, a label collection scheme was implemented to take into account the inter-reader variability of spectral evaluation by multiple MR spectroscopy experts. This is a key step in improving the generalizability of our algorithm, since each reader is independently looking for particular spectral patterns based on his or her own expertise. As shown in Table 0.1**,** there often was disagreement amongst the experts, including approximately 5% of data with complete discordance. As such, the performance of classifiers reflects subjectivity of the human raters that define the ground truth. Multi-center and multi-expert analyses have previously been conducted for brain tumor

classification, and have shown the necessity of such data in establishing quality control norms (83). Inter-rater reliability was assessed by computing the disagreement on spectra for every pair of raters, including the seven human spectroscopy experts and the CNN. The multidimensional scaling plot in Figure 0.4c shows that the distance between the CNN and any of the human raters is not more extreme than the distance between any pair of human reviewers, suggesting that the CNN algorithm agrees with humans about as well as humans agree with each other. Furthermore, the CNN is close to the geometric center of the human raters, which is in accordance with the methodology used to train the CNN; the spatial distance can be attributed to consensus scheme used during data labeling, which deviates from the mean of the three user labels. The OSCAR platform designed for this work is readily available to be used for future sMRI experiments requiring multi-user input. For this work, spectroscopy experts were asked to judge the quality of spectra only based on the metabolite spectrum and its location in a 2D slice. In reality, experts use additional information when assessing spectroscopy data: the strength of the water signal; comparison of the fitted and unfitted spectra; and the pathology of the spectrum's location, e.g. from tumor or healthy tissue. These will be added to OSCAR to supplement data collected in future studies.

Another challenge in developing algorithms for spectral quality filtering is the low percentage of poor quality voxels present in a whole-brain volume compared to good quality voxels, which yields an imbalance in class proportions and consequently can hinder algorithm performance (66). In the dataset collected in this work, 72% of spectra were of good quality and 28% were of poor quality, which is similar to proportions (65 – 84% acceptable spectra) observed in other works (54,55,66). To assess whether balanced class

proportions would affect CNN performance, a random minority oversampling scheme was implemented, wherein data from the minority class (poor spectra) are randomly sampled multiple times to artificially increase the number of samples. A new CNN was trained using the same network architecture on this oversampled dataset, with 5991 good quality and 5991 poor quality, and tuned using Bayesian optimization as described in Methods. The resulting AUC was slightly improved compared to the original CNN, at 0.960 (Figure 0.9). This suggests that the CNN is robust to the class imbalance of the ground truth dataset, but that a more balanced data set could potentially improve outcomes. Of note, the imbalance is relatively small, being only a factor of ~2.6, whereas in other domains where deep learning has been applied the imbalance can be several orders of magnitude (84).

Based on the results of GradCAM, the user can glean some insight into how the network arrives at a decision. An ideal "good" spectrum is simulated by averaging all spectra labeled as "Good", creating a high signal-to-noise ratio spectrum with all peaks visible (Figure 0.7). The GradCAM evaluation of this simulated spectrum suggests that the network appears to focus on regions outside of the main metabolite peaks, specifically on unsuppressed lipids, unsuppressed water, and the overall baseline waveform. Because all spectra evaluated by spectroscopy experts were already passed through a linewidth filter, they were known to have metabolite peaks with narrow linewidths and low Cramer-Rao bounds. As such, the artifacts are arising from other aspects of the spectrum, and the CNN is focusing on these other spectral features. GradCAM visualization also provides a benefit over traditional machine learning methods, in that it can localize the band(s) of an individual spectrum that explain a CNN classification result. In contrast, interpretation of random forests can only be performed on a "population" level, explaining what features

are critical in performing classification in the entire dataset, and cannot provide insights into the classification of individual spectra.

In order to incorporate the spectral quality CNN into a clinical research workflow, a Python application programming interface to query the CNN from other software was developed. The CNN model can perform inference of an entire whole-brain volume on an affordable GPU or on a multicore CPU and has been deployed to an in-house sMRI web app so that it can be easily applied to clinical studies. While training a CNN is computationally intensive and requires powerful GPU hardware, applying the CNN is much less intensive and can be done on commodity hardware, either a low-end GPU or a high-end CPU like the configurations described in the Results section.

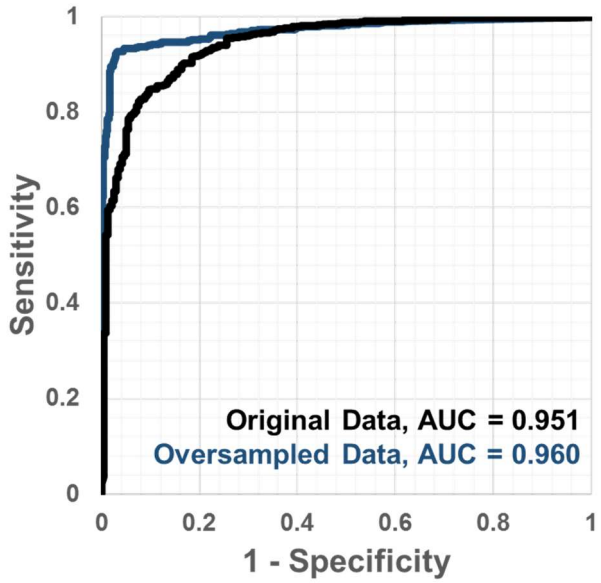


Figure 0.9: Effect of class imbalance on CNN performance. To assess the effect of class imbalance, a data set was generated via oversampling of the poor quality spectra in the original data set, and a CNN was trained on the new data. The oversampled data CNN yielded an improved AUC of 0.960 compared to the original data CNN.

1.13 Conclusion

A deep learning algorithm was developed to automatically detect poor quality spectra in whole-brain sMRI data that otherwise would lead to incorrect classification of voxel pathology. This approach achieved high accuracy when compared to a consensus decision from a panel of MR spectroscopy experts, and achieved comparable accuracy to classifiers based on engineered features developed in this work and by others. The key advantage of the CNN developed here is its ability to operate directly on spectra, enabling quality

filtering independent of the fitting algorithm and requiring no engineered features. The CNN identifies poor quality spectra and artifacts with a high degree of sensitivity and specificity, and has been integrated into a whole-brain spectroscopy processing pipeline. Frameworks such as convolutional neural networks are well suited to the high dimensionality of sMRI data and can extract information that contributes to spectra quality from the spectral waveform beyond what is extracted using engineered features. To provide feedback to spectroscopy experts, we also implemented a GradCAM-based visualization approach that localizes artifacts in spectra, and that can provide a rationale for classification decisions to the user on individual spectra. In this first iteration, the CNN was trained to identify spectral quality based only on spectral waveforms with no regard to other factors that play a role in experts' assessments of quality, such as location of the voxel or any known pathologies in the brain. Future work will focus on collecting training data and developing a CNN that can take these into account, by incorporating an anatomic atlas and assessing local pathology. Ultimately, the implementation of an automated spectral quality filter will mitigate errors due to incorrect classification of pathology, and assist in pushing sMRI into clinical decision-making.

THE BRAIN IMAGING COLLABORATION SUITE

1.14 Author's Contribution and Acknowledgement of Reproduction

This chapter is reproduced from SS Gurbani, BD Weinberg, E Schreibmann, S Sheriff, A Maudsley, MZ Goryawala, LAD Cooper, HK Shu, and H Shim, "The Brain Imaging Collaboration Suite (BrICS): a cloud platform for integrating whole-brain spectroscopic MRI into the radiation therapy planning workflow," *Tomography* (in press). Permission has not been obtained from the publisher at the time of writing of this chapter but will be obtained once the manuscript is available in print.

SSG was the main developer of the software described in this paper, and scanned the subjects and volunteers whose imaging is used in this work. ES, SS, MZG, LADC, AM, and HS assisted with development and validation of several of the software modules. BDW and HKS provided clinical feedback and validated the workflows for the software. HS and HKS are the principal investigators of a clinical pilot study that this software was used for.

1.15 Abstract

Glioblastoma has poor prognosis with inevitable local recurrence despite aggressive treatment with surgery and chemoradiation. Radiation therapy (RT) is typically guided by contrast-enhanced T1-weighted MRI for defining the high dose target and T2-weighted fluid-attenuation inversion recovery (FLAIR) MRI for defining the moderate dose target. There is an urgent need for improved imaging methods to better delineate tumors for focal RT. Spectroscopic magnetic resonance imaging (sMRI) is a quantitative imaging technique that enables whole-brain analysis of endogenous metabolite levels, such as the ratio of choline to N-acetylaspartate (Cho/NAA). Previous work has shown that Cho/NAA accurately identifies tissue with high tumor burden beyond what is seen on standard imaging and can predict regions of metabolic abnormality that are at high risk for recurrence. To facilitate efficient clinical implementation of sMRI for RT planning, we developed the Brain Imaging Collaboration Suite (BrICS, **https://smri.emory.edu**), a cloud platform that integrates sMRI with standard imaging and enables team members from multiple departments and institutions to work together in delineating RT targets. BrICS is being used in a multisite pilot study to assess feasibility and safety of dose-escalated RT based on metabolic abnormalities in patients with glioblastoma (Clinicaltrials.gov NCT03137888). The workflow of analyzing sMRI volumes and preparing RT plans is described. The pipeline achieved a rapid turnaround time by enabling team members to perform their delegated tasks independently in BrICS when their clinical schedules allowed. To date, 18 patients have been treated using targets created in BrICS and no severe toxicities have been observed to date.

1.16 Introduction

The standard-of-care treatment for glioblastoma, the most common adult primary malignant brain tumor, consists of maximal safe surgical resection of tumor followed by high dose radiation therapy (RT) with concomitant temozolomide chemotherapy (TMZ) (85-88). The standard high dose prescription of 60 Gy is delivered over 30 fractions to regions of enhancement on T1-weighted contrast-enhanced (CE-T1w) MRI, in which enhancement represents areas of tumor with leaky neovasculature. A lower dose of RT (typically 46-54 Gy) is delivered to areas of hyperintensity on T2-weighted fluid-attenuation inversion recovery (FLAIR) MRI (9). FLAIR hyperintensity corresponds to a non-specific combination of tumor and non-tumor pathologies, including inflammation and vasogenic edema (89). Despite improvements in maximal resection, concurrent and adjuvant chemotherapy, and RT, the median overall survival still remains poor at 15 months (90,91), with median progression-free survival (PFS) at only 4-6 months (92). Recurrent glioblastoma is very difficult to treat, often being resistant to further radiation and inaccessible for secondary surgical resection (11). The location of recurrent disease can also vary: within the original 60 Gy RT target, within the intermediate dose area, or to regions several centimeters away, including crossing the midline (12). Both local and distant recurrences need to be addressed to improve progression-free-survival. In a phase II study where glioblastomas were treated with high dose proton therapy up to 90 cobalt-gray equivalent (CGE), it was observed that almost all disease recurred in regions receiving less than 70 CGE (93). Thus, it appears that dose escalation may provide sufficient tumoricidal doses to achieve local control. However, doses greater than 70 Gy need to be

applied selectively to prevent toxicity that could result from excess volumes of normal brain being treated to doses of that magnitude.

Spectroscopic magnetic resonance imaging (sMRI) is an evolution of MR spectroscopy (MRS) that enables 3D whole-brain volumes of metabolite levels to be obtained *in vivo* without contrast agents or radioactive tracers (21,94). Two metabolites of particular interest in patients with glioblastoma include choline-containing compounds (Cho), the building blocks of the cell membrane that increase in proliferating tumor cells, and N-acetylaspartate (NAA), a biomarker found in healthy neurons which diminishes due to neuronal displacement and death from glial infiltration (21,62). It has been previously shown via histological correlation that the ratio of Cho to NAA is significantly elevated in glioblastoma due to the opposing changes in these metabolites; in particular, a two-fold increase in Cho/NAA compared to healthy tissue in contralateral normal-appearing white matter (NAWM) was able to correctly identify tumor in 100% of cases, even when tissue samples were biopsied from regions outside of contrast-enhancement per CE-T1w or FLAIR hyperintensity (20).

A combination of dose escalation guided by sMRI, including regions of occult tumor normally left untreated by high dose RT, could potentially delay recurrence of disease by delivering a cytotoxic dose of radiation to regions of metabolically abnormal tumor even if these areas are not detected using standard imaging techniques. However, the use of sMRI in clinical practice has been hampered by data processing requirements and limited integration into the RT planning workflow. In previous studies, several time-intensive manual processing steps were required to import metabolite volumes into clinical imaging software so that they could be used in the operating room or for RT planning (19,95). To

enable integration of sMRI into clinical practice, we have developed a software platform designed specifically for the integration of sMRI into the RT planning workflow. In this paper, we describe its architecture and demonstrate its features on several sample cases. We demonstrate feasibility of this software for collaborative use in a prospective multi-institutional clinical study to target dose escalated RT based on sMRI. Several challenges in integrating this imaging modality into the clinical workflow are addressed, and a sample case from the ongoing study is presented to demonstrate that radiation therapy to high risk regions can be targeted by quantitative imaging techniques such as sMRI.

1.17 Materials and Methods

1.17.1 Software Architecture

To assist with a collaborative clinical study across institutions, we developed the Brain Imaging Collaboration Suite (BrICS), a web-based software designed specifically to integrate sMRI with clinical 3D MRI volumes, enabling physicians to evaluate relevant metabolite levels and the underlying spectra used for this quantitation, and to delineate target volumes for RT planning based on this information (96). BrICS consists of two components: a centralized server and a lightweight browser client (Figure 0.1a). The server performs computations necessary to analyze and display whole-brain spectroscopy; it consists of modules written in C++ and the PHP server-side scripting language to take advantage of well-established image processing and linear algebra libraries data (97,98). The lightweight browser client is written in Javascript and can run on all modern hardware, including thin clients such as laptops and tablets. This cloud approach offers several

benefits over standalone software clients: 1) improves repeatability and standardization by ensuring data are processed on the same hardware; 2) reduces user variability and bias; 3) enables real-time deployment of software updates across all clients; 4) prevents the need for every end-user to download massive sMRI data sets onto a local computer; 5) runs without the need for the user to download any software beyond a web browser, which is of key importance since physicians often use restricted hospital workstations; and 6) allows information and images to be easily shared with patients who wish to be better informed of their clinical management.

BrICS imports data from spectroscopy processing software, such as the Metabolite Imaging and Data Analysis Software (MIDAS), and from other imaging systems/software using the Digital Imaging and Communication in Medicine (DICOM) file format. All volumes are co-registered using a rigid transformation and resampled using trilinear interpolation into a high-resolution T1w image space, enabling overlays of metabolic information onto anatomic MRI. Users can then delineate target volumes based on both anatomic and spectroscopic information. These targets can be exported as DICOM radiation therapy structure sets (DICOM RT) or binary DICOM masks and imported into RT planning systems to deliver therapy to patients (Figure 0.1b).

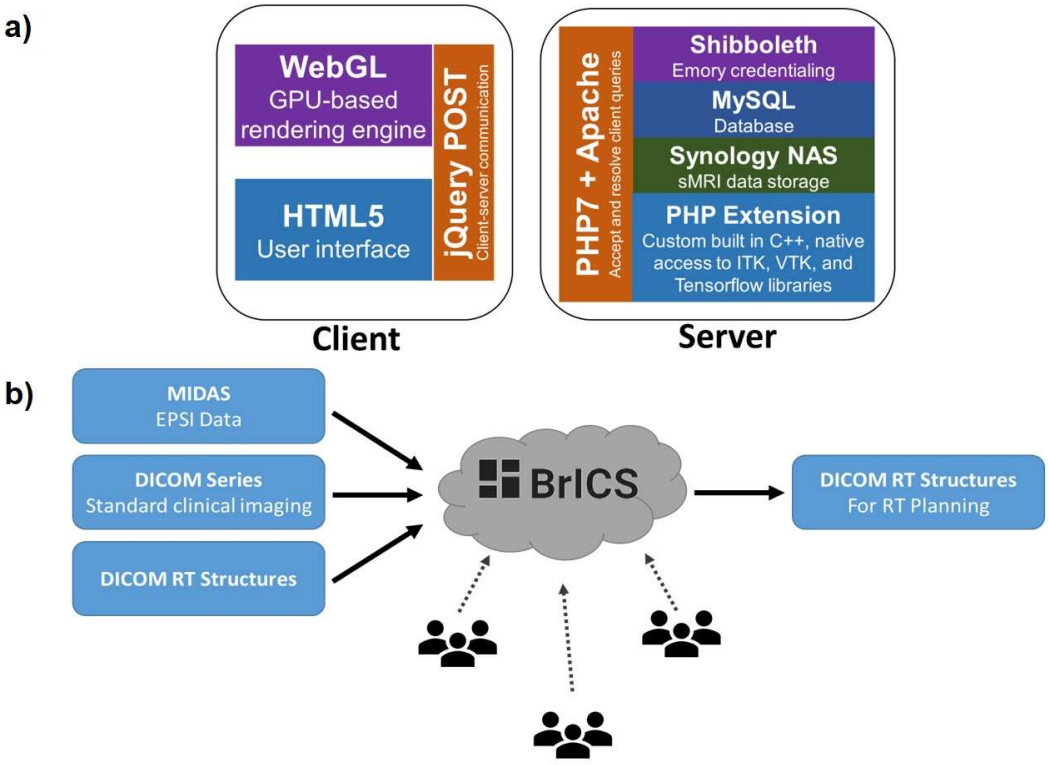


Figure 0.1: Schema of BrICS architecture. a) The Brain Imaging Collaboration Suite (BrICS) consists of a centralized server which performs image processing, and a lightweight browser client. b) BrICS imports spectroscopy and DICOM-format MRI volumes, and can export RT targets to other clinical software. sMRI volumes are blended with clinical MRI, and users can perform tasks such as evaluating underlying spectra and contouring based on sMRI abnormalities.

1.17.2 Visualization and Contouring

The main interface of BrICS is shown in Figure 0.2. sMRI volumes – either individual metabolites or metabolite ratios – are overlaid on anatomic volumes (e.g. T1w MRI), enabling visual assessment of metabolic changes in spatially dependent manner. For MR spectroscopists and radiologists familiar with MRS techniques, selection of a voxel will bring up the corresponding spectrum. Since sMRI is a quantitative imaging technique, voxel intensities can be reliably interpreted across subjects, and decision-making can be based on specified thresholds. This ability is built-in to the contouring module; physicians can make contours based on the values in sMRI maps (Figure 0.3). For example, the Cho/NAA volume abnormality index (20) can be used, as shown, to generate a contour around all voxels which have a Cho/NAA abnormality index above a given threshold; users can select this threshold and automatically generate contours of increasing or decreasing sensitivity of disease detection. Radiologists can then review these contours and make changes to them using built-in editing tools (painting, erasing, selection of connected-components). Once contours are generated, they can be visualized as 3D objects, enabling visual quality assessment and correspondence with anatomy. Statistics such as contour volume and number of connected components are also reported.

1.17.3 Normalization of Metabolite Values

Cerebral concentrations of several macromolecules, including Cho and NAA, are known to vary based on a subject's age, gender, and anatomic location of brain being sampled (17). To account for these variations in baseline metabolism, metrics such as the Cho/NAA abnormality index (20) and the Cho-NAA-index (42) take into account relative

changes in these metabolites compared to normal tissue, typically contralateral normal-appearing white matter (NAWM) (99). For this trial, we use the Cho/NAA abnormality index, defined as the Cho/NAA of a given voxel divided by the mean Cho/NAA value in contralateral NAWM. In previous works (19,20,95), NAWM was manually contoured on a clinical T1w volume by a neuroradiologist using commercial software, then the mask exported and applied to sMRI data to determine the mean Cho/NAA value.

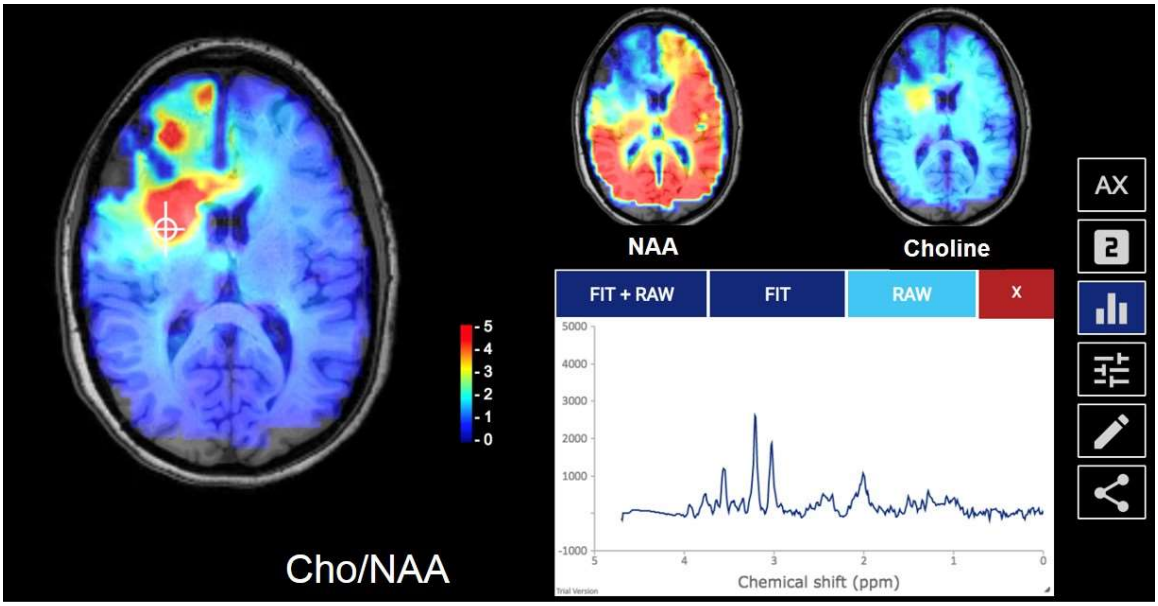
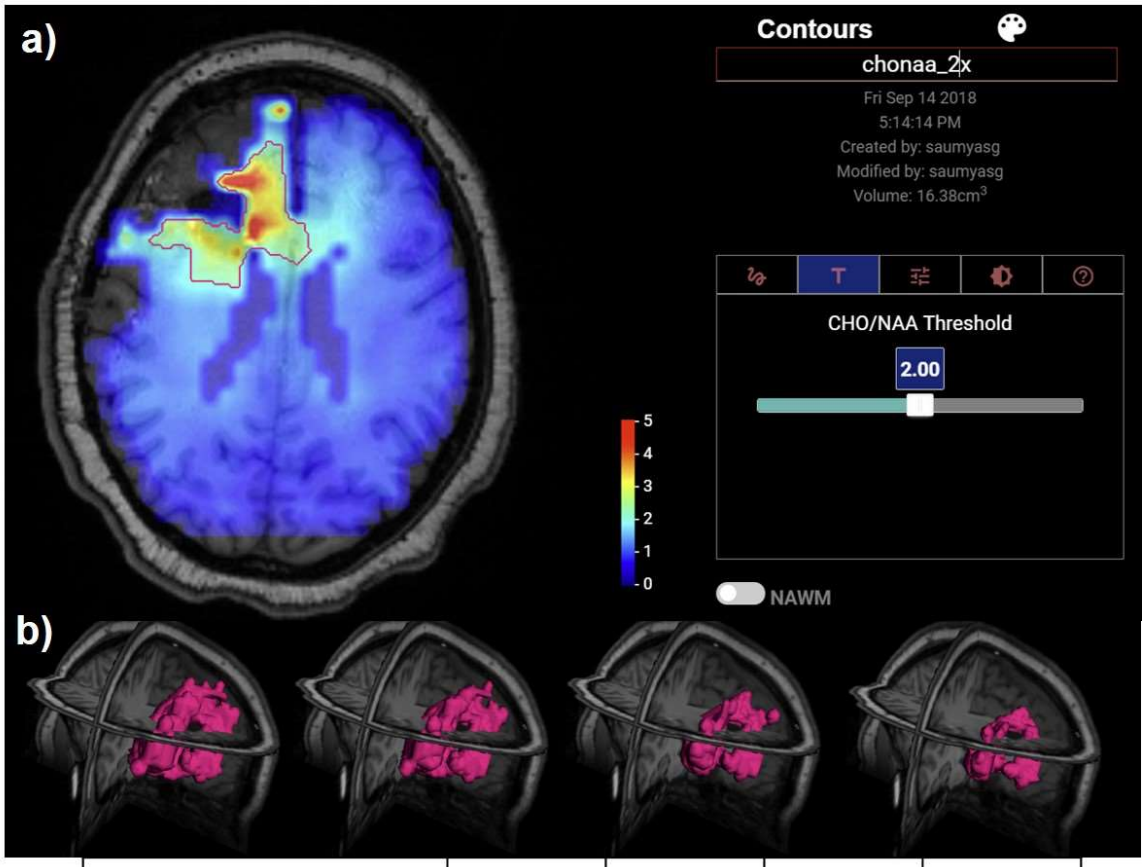


Figure 0.2: The main user interface for BrICS. sMRI metabolite and metabolite ratio maps are overlaid on top of anatomic MR volumes. Selection of a given voxel brings up the underlying spectrum.



c)

Cho/NAA Abnormality Index	1.75X	2.0X	2.5X	3.0X
Contour Volume (cc)	63.98	48.11	29.19	16.38

Figure 0.3: The BrICS contouring module. a) The contouring module enables identification of target volumes based on either anatomic or metabolite images. For quantitative imaging techniques like sMRI, users can automatically delineate contours using threshold values; b) A series of targets based on thresholding of the Cho/NAA abnormality index; target volumes can be rendered in 3D for visual inspection prior to being exported to other clinical software.

To expedite this process, remove reliance on commercial software, and mitigate user bias, we have implemented an algorithm in BrICS to automatically contour the NAWM based on a Gaussian mixture model (100) (Figure 0.4). First, all voxels from the cerebrum are masked using an anatomic atlas. Next, all cerebral Cho/NAA voxels are modeled as a bimodal Gaussian distribution, with voxels arising from the second, higher-mean Gaussian population representative of tumor pathology. These voxels are then masked, and the side with largest contiguous abnormal segment is selected as the side of tumor; voxels in the contralateral hemisphere are segmented into gray and white matter based on fractional water content (101) calculated by MIDAS, and then the mean is reported as the normalizing factor for the subject's abnormality index calculations.

1.17.4 Automated Segmentation of Residual Contrast Enhancement

Additional algorithmic modules can be built into BrICS to assist with other routines that are regularly performed by clinicians. One such module automatically contours residual contrast enhancing tissue (Figure 0.5), so as to differentiate true unresected tumor with leaky neovasculature from blood products due to surgical resection (102). The module requires a pre-contrast T1w MRI, a CE-T1w MRI, and a T2w or FLAIR MRI, all of which are co-registered into the same imaging space in the axial plane. The pre- and post-contrast MR images are histogram normalized and subtracted to generate a difference map; Otsu thresholding with four classes is used to identify residual enhancement (103,104). Otsu thresholding is applied to the FLAIR map to automatically identify hyperintensity; the single largest connected component is used as a bounding mask for the T1w residual

volume. Finally, morphological opening and closing filters are applied to the bounded T1w residual volume to remove islets and thin anisotropic components, e.g. blood vessels. The entire algorithm can be run in < 10 seconds on the BrICS server and yields a final contour which can be evaluated and manually edited, if necessary, by a neuroradiologist - saving valuable clinician time and providing a reproducible starting point for all users.

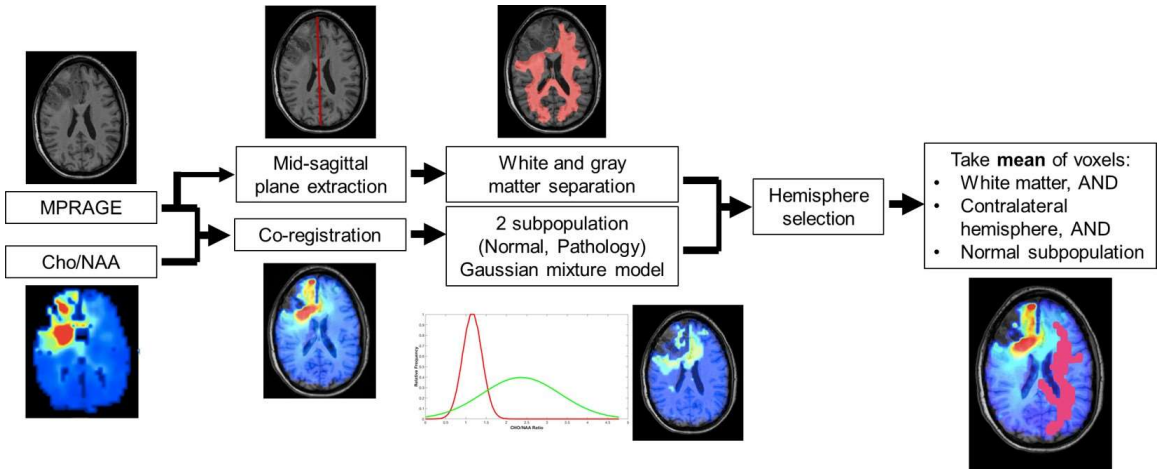


Figure 0.4: Schema of the Gaussian mixture model for NAWM segmentation. NAWM is typically contoured manually by radiologists; this algorithm can perform the same contouring automatically in just a few seconds.

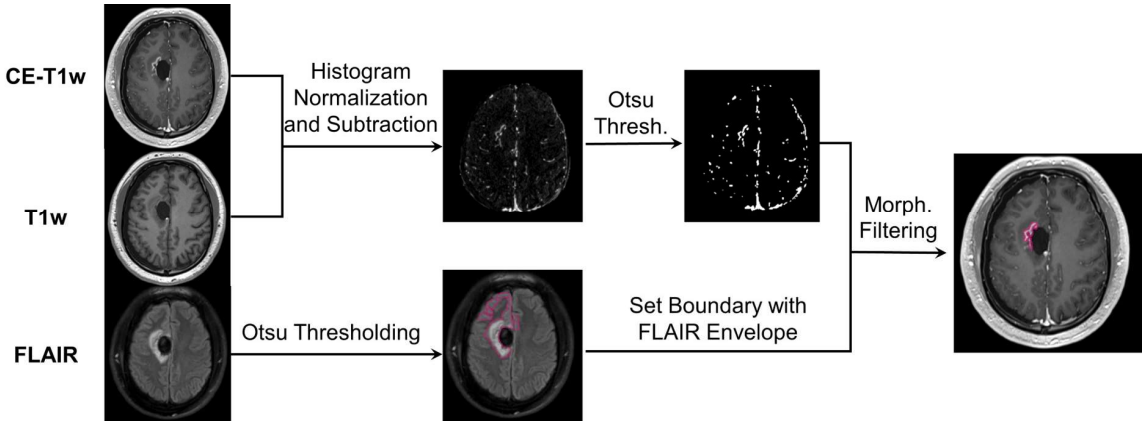


Figure 0.5: Algorithm for automated residual contrast enhancement segmentation. BrICS takes a post-contrast T1w MRI (top), pre-contrast T1w MRI (middle), and a FLAIR MRI (bottom) volume, and follows the shown algorithm to rapidly contour residual contrast enhancement post-surgical resection. This volume can then be edited manually by the neuroradiologist or radiation oncologist as desired to define a dose-escalated volume.

1.17.5 Patient Imaging

To assess the feasibility and safety of sMRI-guided radiation therapy, a multisite clinical study funded by NCI was initiated (Clinicaltrials.gov NCT03137888). Three institutions are participating in this pilot study – Emory University, the Johns Hopkins University, and the University of Miami – and a total of 30 patients with newly diagnosed glioblastoma will be enrolled. Patients enrolled after undergoing maximal safe surgical resection or biopsy at the discretion of the neurosurgeon. Enrolled patients were ≥18 years of age, had a Karnofsky Performance Score ≥ 60, and were willing to undergo dose-escalated radiation therapy to 75 Gy.

An sMRI scan was obtained up to two weeks prior to starting RT+TMZ. A 15 minute echo planar spectroscopic imaging (EPSI) pulse sequence combined with GRAPPA (parallel imaging (18)), was performed on a 3T scanner (Siemens Medical Solutions, Erlangen, Germany) with a 32-channel or 20-channel head coil array (echo time [TE] = 50ms, repetition time [TR] = 1551ms, flip angle [FA] = 71 degrees). During the same session, a high resolution T1w magnetization prepared rapid acquisition gradient echo (MP-RAGE) sequence was obtained at the same orientation and position as the EPSI. Raw EPSI data was transformed into spatial-spectral data, co-registered with the MP-RAGE volume, and the relative metabolite concentration values were obtained by spectral fitting using MIDAS (17,18).

1.17.6 Radiation Therapy Planning

An outline of the workflow for patients in this study is shown in Figure 0.6. The EPSI/GRAPPA and MP-RAGE volumes, in addition to the most recent clinical CE-T1w and FLAIR MRI, were imported into BrICS. Automated contours for Cho/NAA abnormality index of 2.0 and residual contrast enhancing tissue were generated using the algorithms described above. Using BrICS, two MR spectroscopists from different institutions collaboratively reviewed the underlying raw and fitted spectra within the Cho/NAA abnormal contour and removed voxels with poor spectral quality. Meanwhile, a neuroradiologist then reviewed and edited the residual contrast enhancing volume to ensure the algorithms accurately captured the target volumes, and validated the residual CE volume. The two separate contours were then merged to form a single target volume for

high dose RT. Next, an external radiation oncologist (from a non-treating site) edited and approved the volume based on anatomy and dose safety concerns. Finally, the treating-site radiation oncologist made final edits based on his/her discretion and validated the volume for RT treatment. To ensure patient safety and to enable retrospective review of this study, all user edits were tracked in BrICS in a digital audit trail.

The final contour generated in BrICS was defined as gross tumor volume 3 (GTV3). The clinical target volume 3 (CTV3) was defined as equal to GTV3 with no margin. In this pilot feasibility study, a maximum volume of 65cm^3 was allowed for CTV3, approximately adhering to the 5 cm diameter boost volume limit used in the NRG Oncology BN001 phase II trial on RT dose escalation for glioblastomas (105). The CTV3 contour was exported from BrICS as a DICOM RT structure set on the high resolution T1w MP-RAGE volume into additional contouring or treatment planning software such as VelocityAI (Varian Medical Systems), MIM Maestro (MIM Software Inc), Eclipse (Varian), Pinnacle (Philips Healthcare), etc., per the routine of the treating site. Additional standard treatment volumes were generated including GTV2, defined as the surgical cavity with residual contrast-enhancement, and GTV1, defined as hyperintensity on FLAIR MRI. 5mm anatomically-constrained margins were added to GTV2 and GTV1 to produce CTV2 and CTV1, respectively. A 3mm margin was added to all three CTVs to produce the planning target volumes (PTV3, PTV2, and PTV1). A simultaneous in-field boost IMRT plan was generated to treat PTV3, PTV2, and PTV1 to 75 Gy, 60 Gy, and 51 Gy, respectively, respecting standard organs-at-risk constraints (Table 0.1).

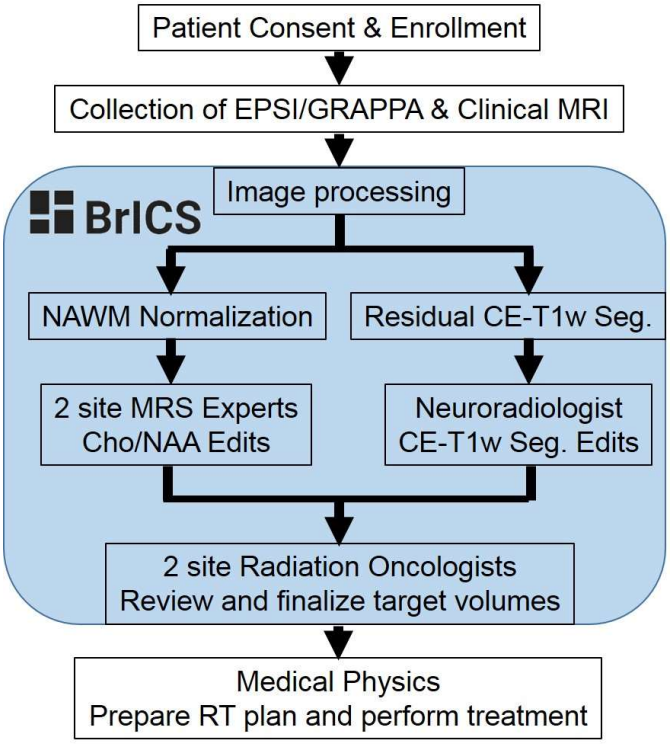


Figure 0.6: Radiation therapy planning workflow for sMRI-based dose escalation. After patients are enrolled and consented, their imaging data are processed and edited in BrICS. The centralized platform enables reliable and repeatable processing, with documented edits made by physicians and spectroscopists to prepare the final treatment plan.

Table 0.1: Summary of target volume definitions and dose prescriptions.

Target Name	Definition	CTV Margin	PTV Margin	Dose
GTV3	Cho/NAA abnormality index ≥ 2 + residual contrast enhancement	0mm	3mm	75 Gy
GTV2 (standard of care)	Contrast enhancing tissue + resection cavity	5mm	3mm	60 Gy
GTV1 (standard of care)	FLAIR hyperintensity	5mm	3mm	50.1 Gy

1.18 Results

A demo of BrICS is available at **https://brainimaging.emory.edu/brics-demo** with a few curated and de-identified data sets. In addition to the dose-escalated RT study described above, BrICS is currently being used for several clinical projects internally at Emory University: targeting of biopsies in patients with non-enhancing low-grade gliomas, monitoring therapeutic response of patients with glioblastoma receiving a histone-deacetylase inhibitor in addition to standard chemoradiation, identification of metabolite abnormalities associated with melanoma brain metastasis, and a pilot study evaluating the benefit of sMRI for patients with mild traumatic brain injury. Additionally, BrICS served as the platform for testing new imaging processing algorithms such as a neural network for identifying spectral artifacts (106) and autoencoder-based spectral fitting (107).

RT plans from one patient who underwent dose escalation per this study's protocol are presented in Figure 0.7. The patient is a 21-year-old female diagnosed with a frontal glioblastoma and enrolled in the trial one month after undergoing surgical resection of her tumor. sMRI volumes were obtained and processed in MIDAS and in BrICS per the protocol. The neuroradiologists, spectroscopists, and radiation oncologists accessed BrICS remotely for several minutes each, when time was available during their busy schedules. Segmentation of residual contrast enhancement by the automated algorithm, followed by neuroradiologist review, identified a 1.6 cm3 nodular residual contrast-enhancing lesion on the posterior border of the surgical cavity remaining after surgery, typical of patients who underwent near total resections (Figure 0.7a). However, the Cho/NAA abnormality was much greater with a volume of 50.6 cm3, expanding laterally, anteriorly, and posteriorly from the surgical cavity (Figure 0.7b). GTV3 was planned on the union of these

two contours and targeted for a 75 Gy boost. The contour for GTV3 was exported as a DICOM RT structure and imported into Eclipse for dose planning (Figure 0.7c). Dose constraints based on RTOG guidelines to all organs-at-risk (OARs) were met, with greater than 95% of the prescribed dose delivered to each PTV.

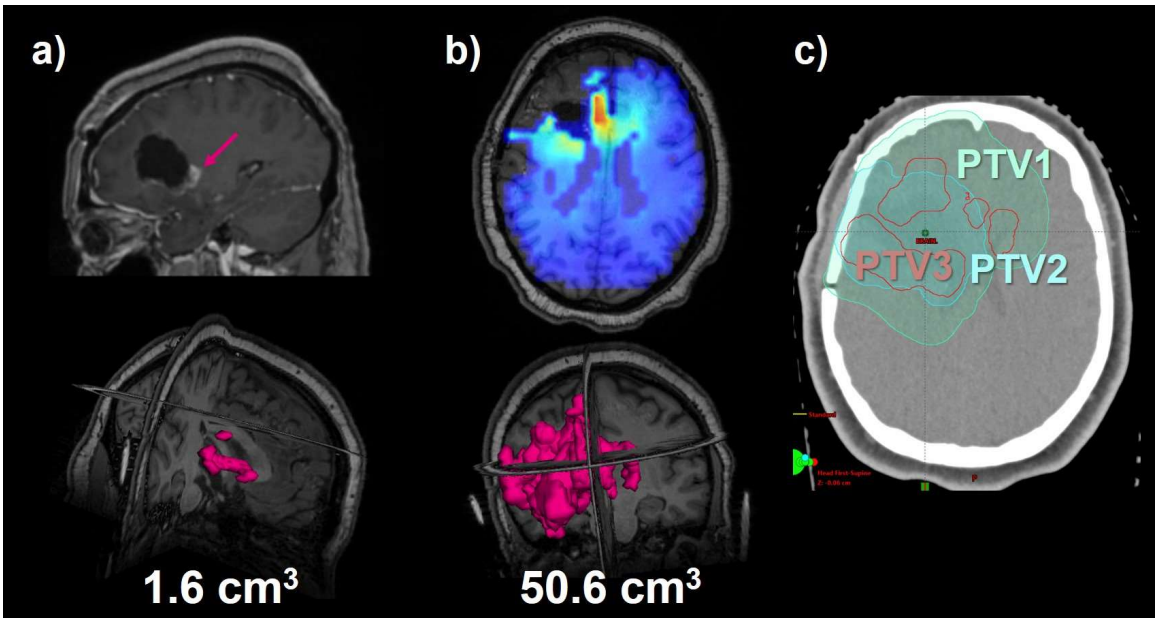


Figure 0.7: Example treatment plan for study patient. a) The patient is a 21-year-old female with newly diagnosed glioblastoma with a near-total resection of the tumor. b) However, the Cho/NAA map indicates metabolically active tumor expanding outward from the resection cavity. c) A boosted dose of 75 Gy (PTV3) was successfully planned and delivered to this patient.

1.19 Discussion

Current treatments for glioblastoma are insufficient in achieving local control. This is felt to be due in part to limitations of standard imaging methods in identification of infiltrating tumor margins, which show no contrast-enhancement, potentially leaving these high-risk regions undertreated. Improvements in treatment options, such as with higher radiation doses, can only be beneficial if all high-risk tumor regions (both enhancing and non-enhancing) are properly targeted. In this work, we develop a software platform that successfully enables sMRI integration into the radiation therapy planning workflow. The EPSI/GRAPPA sequence can be used on standard 3T instrumentation, and the current version of the sequence is available for several different Siemens models (e.g. PRISMA, Trio, and Skyra); expansion to other vendors is an ongoing project. The data can then be sent to a centralized server for processing, Because it is web-based, BrICS can be used by multiple users and institutions without the need for additional software, data, or processing. BrICS was successfully used as the infrastructure for an ongoing multi-institutional clinical study assessing the feasibility of dose-escalated radiation guided by sMRI in patients with glioblastoma; to date, 18 patients have been treated on this protocol, and no toxicities have been observed. Thus, there is an urgent need for improved quantitative imaging biomarkers that can not only identify these regions, but also be readily incorporated into clinical practice.

sMRI has been shown to delineate infiltrating tumor beyond standard MRI but has thus far been only used in retrospective analyses due to the complexity of integrating it with clinical volumes, the requirement for an on-site MR spectroscopist to manually review spectra in individual voxel, and variability in acquisition and processing across institutions.

A web-platform such as BrICS provides solutions for these challenges by enabling centralized data storage and analysis, allowing clinicians from multiple institutions to utilize sMRI without the need for local experts or software. Users will always have the latest version of BrICS without needing to download additional software or data sets and can access BrICS from any computer browser. BrICS is currently being used in a multisite clinical study assessing the feasibility of sMRI guidance for radiation therapy and will continue to be developed as infrastructure for future consortium-level trials.

1.20 Acknowledgements

The authors would like to thank the following people for their assistance in collecting and analyzing data: Dr. Peter Barker, Dr. Michal Povazan, Dr. Sarah Dupont, Mr. Michael Larche, Mr. Robert Smith, Ms. Samira Yeboah, and Ms. Sarah Basadre. This work is funded by NIH grants U01CA172027, R01CA214557, U01EB028145, and F30CA206291.

1.21 Author's Contribution and Acknowledgement of Reproduction

This chapter is reproduced from SS Gurbani, S Sheriff, AA Maudsley, H Shim, and LAD Cooper. *Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting*. Magn Res Med, 2019; 00: 1-12. Permission has been obtained from the publisher, John Wiley and Sons. License number 4514210041428.

SSG designed the convolutional encoder-model decoder architecture. SSG and SS performed data collection. All authors performed data analysis and assisted with manuscript preparation.

1.22 Abstract

1.22.1 Purpose

Magnetic resonance spectroscopy imaging (MRSI) has shown great promise in the detection and monitoring of neurologic pathologies such as tumor. A necessary component of data processing includes the quantitation of each metabolite, typically done through fitting a model of the spectrum to the data. For high-resolution volumetric MRSI of the brain, which may have ~10,000 spectra, significant processing time is required for spectral analysis and generation of metabolite maps.

1.22.2 Methods

A novel deep learning architecture that combines a convolutional neural network with a priori models of the spectrum is presented. This architecture, a convolutional encoder – model decoder (CEMD), combines the strengths of adaptive and unbiased convolutional networks with models of magnetic resonance and is readily interpretable.

1.22.3 Results

The CEMD architecture performs accurate spectral fitting for volumetric MRSI in patients with glioblastoma, provides whole-brain fitting in one minute on a standard computer, and handles a variety of spectral artifacts.

1.22.4 Conclusion

A new architecture combining physics domain knowledge with convolutional neural networks has been developed and is able to perform rapid spectral fitting of whole-brain data. Rapid processing is a critical step toward routine clinical practice.

1.23 Introduction

Proton magnetic resonance spectroscopic imaging (MRSI) is an imaging modality capable of generating high-resolution 3D maps of cerebral metabolites concentrations in vivo (20,22,62). Previous studies have shown that altered metabolism identified regions of occult tumor that are not visible in contrast-enhanced T1w MRI (20,108,109). Studies have assessed the potential of using MRSI to guide radiation therapy (RT) in patients with glioblastomas (95,110-112), and an ongoing multisite clinical study is prospectively boosting radiation based on this technique (NCT03137888, "Spectroscopic MRI-Guided Radiation Therapy Planning in Glioblastoma") (113). To utilize volumetric MRSI for clinical studies, maps of the individual metabolite distributions must be created by quantifying the metabolite resonance peaks, a process known as spectral fitting. Several spectral fitting algorithms have been established, including ones that operate in both the time and frequency domain of the acquired data (114,115). Common amongst these techniques is the incorporation of prior knowledge, which includes information on the resonance frequencies of metabolites of interest and lineshapes. Several parametric spectral analysis methods have developed (26,28-31,116,117), all of which rely on iterative optimization procedures to find the model parameters that best match the data. However, these methods do not scale well to volumetric spectroscopic imaging, which can contain on the order of 10,000 spectra in a whole-brain scan. One method for processing of whole-brain MRSI is the FITT program of the Metabolite Imaging and Data Analysis System (MIDAS) (17,18), which uses iterative time-frequency parametric modelling methods for spectral analysis (30). This algorithm is typical of the parametric spectral analysis approaches that have been widely applied to MRS. As with other parametric spectral

analysis methods, spectral fitting with MIDAS is computationally-intensive, requiring 40-50 minutes on a high-end multicore workstation. However, if MRSI is to be integrated into clinical protocols it is critical that all processing be done within a few minutes, ideally on board the scanner computer. This would enable rapid quality checks of the final maps to see if the scan needs to be repeated before the patient leaves and to be sent for clinical review in a timely manner.

Machine learning has proven to have exceptional utility in medical imaging, including MRSI (54,55,58,59,118,119). Hiltunen et. al. described an artificial neural network (ANN) architecture that could predict metabolite peak areas from magnitude spectra in patients with gliomas (58); Das et. al. presented a multi-layer perceptron (MLP) for quantifying metabolite concentrations from synthetically generated spectra and phase-encoded 2D MRSI using results from LCModel as data for supervised training, achieving accurate predictions of metabolite concentrations (61); and Bhat et. al. used an unsupervised neural network for analysis of phase-corrected 2D MRSI data (59). While these methods provide measures of metabolite concentrations, it is difficult to contextualize why the MLP or ANN produced a given output. A concept in machine learning that correlates well to curve fitting is the idea of the encoder-decoder, or autoencoder. Autoencoders are a type of unsupervised learning neural network that seek to find a compressed encoding of input data such that the input data can be accurately reconstructed from this parsimonious encoding (120,121). The autoencoder consists of two independent neural networks: an encoder, that transforms the input data into a lower dimensional space; and a decoder, that reconstructs the original input data from the lower dimensional representation (Figure 0.1**Error! Reference source not found.**). The encoder and decoder

are concurrently trained to minimize reconstruction error, with the constraint that the interim representation has significantly lower dimension than the input, and thus only important "features" of the data are maintained. This process is highly analogous to parameterized curve fitting, wherein a model (the decoder) uses relatively few parameters to reconstruct a denoised version of raw data. With autoencoders, both the encoding and decoding models are neural networks (a cascade of convolution, pooling, matrix multiplication and summation operations) such that higher-order non-linear features can be extracted from the inputs. The model parameters of these neural networks can be optimized using gradient descent techniques to minimize reconstruction error. Generally, the only constraint on the autoencoder is the size of the lower dimensional representation, and as such, this low-rank representation is not readily interpretable.

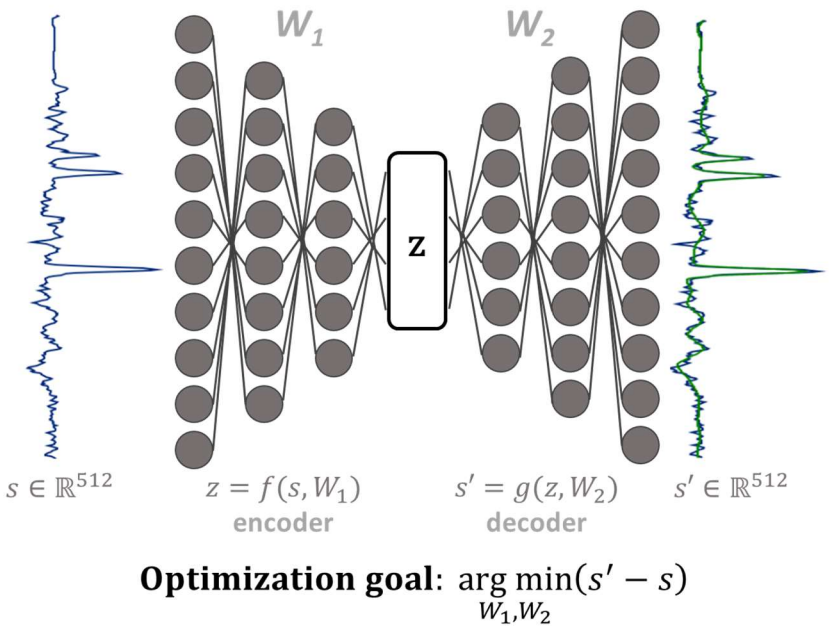


Figure 0.1: Schema of a general autoencoder. An autoencoder takes an input, s, and seeks to encode it into a compressed latent space, z, such that the input can be reconstructed using only information from this space. Both the encoder and decoder are convolutional neural networks.

To leverage the feature-learning capabilities of autoencoders while maintaining the interpretability of parameters given domain knowledge about MRS, a novel spectral fitting algorithm was developed that utilized a convolutional neural network encoder with a model-based decoding of the spectrum. We evaluated this method for fitting of singlet resonances in MRSI of the brain. We also implemented a software pipeline for rapid generation of metabolite maps of interest. Metabolite maps generated using this system were then compared with those produced by existing parametric spectral fitting methods.

1.24 Methods

1.24.1 Image Acquisition and Processing

Volumetric echo planar spectroscopic imaging (EPSI) scans were performed on four healthy subjects and six subjects with newly diagnosed glioblastoma who were enrolled in an ongoing clinical trial (NCT03137888). Scans were conducted at 3T (Siemens Prisma) with a 32-channel head coil (Siemens Healthineers, Erlangen, Germany). For the subjects with glioblastoma, data were obtained following surgical resection but prior to radiation therapy and chemotherapy, as previously described (119). Briefly, T1-weighted (T1w) magnetization-prepared rapid gradient echo pulse (TR = 1900 ms, TE = 3.52 ms, 256 x 256 x 160 matrix, flip angle (FA) = 9°) and whole-brain 3D EPSI (TR = 1551 ms, TE = 50 ms, 64 x 64 x 32 matrix, FA = 71°) sequences with generalized autocalibrating partially parallel acquisitions (GRAPPA) acceleration were obtained during the same scanning session, oriented at a +15 degree tilt in the sagittal plane from the anterior commissure-posterior commissure line (122). For the EPSI/GRAPPA sequence, an oblique saturation band was placed in the sagittal plane from the optic chiasm to the cerebellum. A medium TE of 50ms was used to reduce the impact of lipid contamination and emphasize the signal of metabolites of interest in subjects with brain tumors: choline (Cho), creatine (Cr), and N-acetylasparate (NAA). MIDAS (17,18) was used to perform the following preprocessing and volume reconstruction steps: spatial reconstruction, B0 field correction using simultaneously-acquired intracellular water signal, co-registration of the T1w and metabolite volumes, lipid suppression, and water suppression. The initial processing also created a mask covering the brain volume, but excluding voxels for which the water linewidth was greater than 18 Hz, as calculated using the T2* map. A total of 102,005

spectra were obtained and separated into three data subsets: 85,661 for training; 8,192 for validation; and 8,192 for testing.

For assessing generalizability, additional data from subjects with newly-diagnosed glioblastoma that were scanned at Emory University, the University of Miami and the Johns Hopkins University, were evaluated. The same sequences and parameters as above were used, with the exception that studies at the University of Miami were carried out on a 3T Siemens Skyra with a 20-channel head coil.

1.24.2 Convolutional Encoder – Model Decoder

The goal of spectral fitting for an input spectrum, $s \in \mathbb{R}^{512}$, is to find a parameter set, $\theta \in \mathbb{R}^{k}, k \ll 512$, such that when these parameters are used in a spectral model, $g(\theta)$, the result is a noise-free approximation of the input:

$$g(\theta): \mathbb{R}^{k} \rightarrow \mathbb{R}^{512}, g(\theta) \approx s \tag{1}$$

Mathematically, this can be represented as the following optimization task:

$$\arg\min_{\theta}[g(\theta) - s] \tag{2}$$

In this context, the task of finding θ is the encoding step while the function g(θ) is the decoding step. In this work, we define the convolutional encoder – model decoder (CEMD) architecture. The CEMD architecture is graphically depicted in Figure 0.2. The encoding step in CEMD is a convolutional neural network (CNN) composed of sequential layers of

convolution, pooling, and rectification (73,123). Each of these layers is has parameterized weights that can collectively be referred to as W; therefore, the CNN with weights W can be conceived of as applying a series of transformations that perform the encoding function $f : \mathbb{R}^{512} \rightarrow \mathbb{R}^{k}$ defined by $f(s, W) = \theta$. Training identifies W to minimize the error defined in Equation 2. In the context of spectral fitting, a fixed spectral model is used as the decoder (Equation 1) to generate a fitted version of the input spectrum. The overall training objective can be stated as:

$$\arg\min_{\boldsymbol{W}} \left[g\big(f(s, W)\big) - s \right] \tag{3}$$

A key point in this optimization is that it requires no "ground truth" or "true" fitted spectra for CNN training. It is an unsupervised learning task requiring only a set of spectra, $S = \{s_1, s_2, \ldots, s_N\} \in \mathbb{R}^{512 \times N}$, to train the CNN weights with the goal of minimizing the residual error between the input data and fitted spectra.

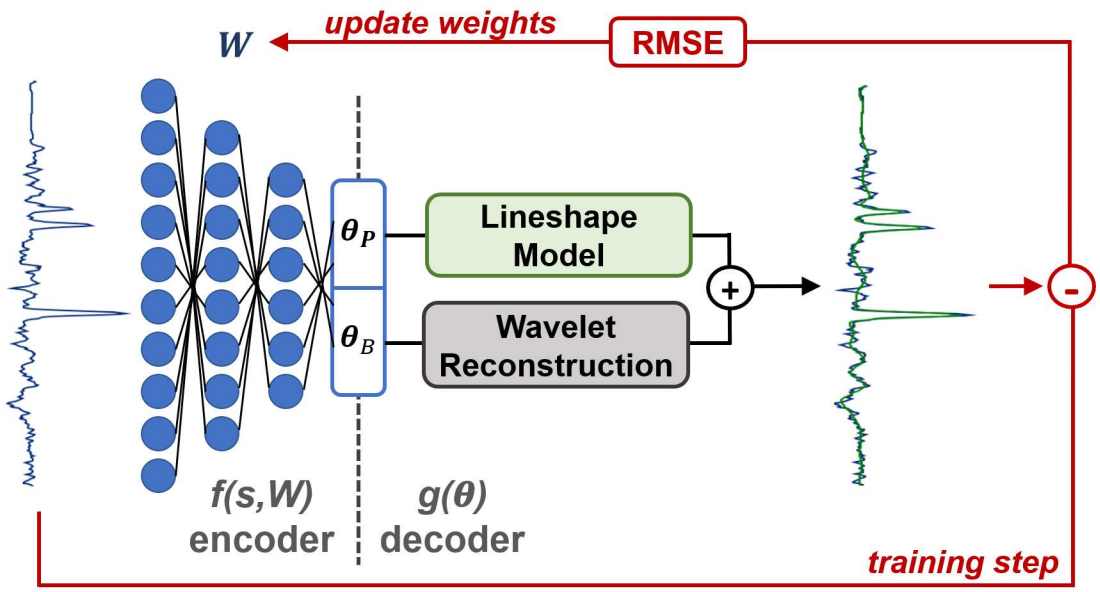


Figure 0.2: Schematic of the convolutional encoder – model decoder (CEMD) architecture. CEMD uses a convolutional neural network (CNN) to learn a low-rank representation of input spectra and uses known models for the baseline and peak components of spectra to train the CNN.

The decoder was based on previously described parametric analyses that model a spectrum as having two components: i) metabolite resonances that are explicitly defined, and ii) a baseline composed of all metabolites and macromolecules not explicitly defined (29,30). Metabolite resonances were modeled using the Lorentzian-Gaussian lineshape model:

$$s = FFT\left(\sum_{m=1}^{3} A_m\, e^{-i(\omega_{m,0}+\Delta\omega_m+\phi_1)t+\phi_0}\, e^{-\left[\frac{t}{T_a}+\left(\frac{t}{T_b}\right)^2\right]}\right) \quad (4)$$

For each metabolite, *m*, the metabolite model required six parameters: peak amplitude A_m, resonance frequency ω_m, zero and first order phases (ϕ_0 and ϕ_1), and Gaussian and Lorentzian decay constants (T_a and T_b). The three major singlet resonances at TE = 50 ms were modeled: Cho, Cr, and NAA. Note that this formulation returns the relative concentrations of each metabolite, not the peak area. A constraint was placed such that all three metabolites have the same zero- and first-order phase shifts and linewidths, such that only the resonance frequency and amplitude needed to be independently determined. Since the expected resonance frequency, defined as $\omega_{m,0}$, is known a priori from a library of chemical resonances, only a shift in frequency from the expected, $\Delta\omega_m$, needed to be calculated. Thus, for these three metabolite singlet resonances, a total of 10 parameters were needed:

$$\theta_P = \{A_{Cho}, A_{Cr}, A_{NAA}, \Delta\omega_{Cho}, \Delta\omega_{Cr}, \Delta\omega_{NAA}, \phi_0, \phi_1, T_a, T_b\} \quad (5)$$

The baseline component was defined by wavelet reconstruction, using a set of coarse (124) third-order Coiflets (125) as the wavelet kernels and four levels of dyadic upsampling

to convert 32 coarse coefficients (θ_B) into the baseline signal. In order to enable automated computation of gradients for training the CNN in TensorFlow (126), wavelet reconstruction was implemented as a series of linear matrix operations. At each level, the output of the previous iteration, $y_{low} \in \mathbb{R}^p$, were first dyadically upsampled (127). Dyadic upsampling on a vector y was implemented as augmentation of its transpose with a vector of zeros of the same size, followed by vectorization and transposition:

$$y = [y_1 \quad \cdots \quad y_p] \tag{6}$$

$$y^{aug} = \left(y^T \middle| \vec{0}_p\right)^T = \left(\begin{bmatrix} y_1 & 0 \\ \vdots & \vdots \\ y_p & 0 \end{bmatrix}\right)^T = \begin{bmatrix} y_1 & \cdots & y_P \\ 0 & \cdots & 0 \end{bmatrix} \tag{7}$$

$$y^{up} = vec(y^{aug})^T = [y_1 \quad 0 \quad y_2 \quad 0 \quad \cdots \quad y_p \quad 0] \tag{8}$$

The upsampled vector was then convolved with the Coiflet low-frequency reconstruction kernel, $Coif3_{low}$ (125):

$$y'_{low} = y_{low}^{up} * Coif3_{low} \tag{9}$$

The central *2p* elements of y'_{low} were used as the input for the next level of wavelet reconstruction. The CEMD encoder therefore needed to calculate 42 coefficients, 10 for the metabolite resonances and 32 for the baseline, which are passed to the decoder to create the fitted spectrum.

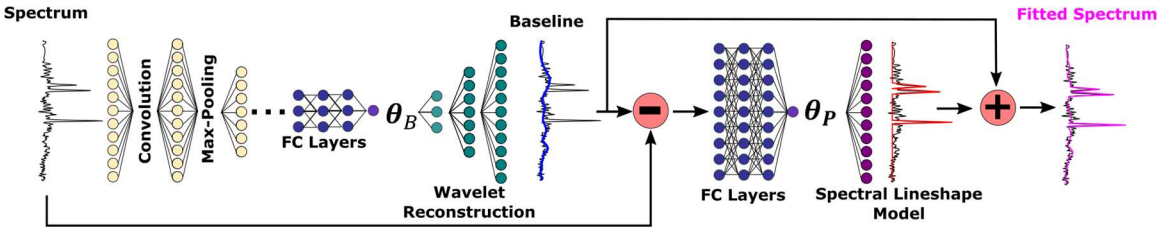


Figure 0.3: Detailed schematic of the CEMD architecture. CEMD features two serial encoder-decoder steps, the first which computes the baseline, and the second which fits resonance peaks on the baseline-subtracted spectrum. FC = fully connected.

A more detailed representation of the CEMD architecture is shown in Figure 0.3. Only the real component of the complex input spectrum ($s \in \mathbb{R}^{512}$) was passed through a CNN during the encoder to produce a low-rank representation that directly corresponded to the 42 spectral model parameters. The decoder then applied the metabolite resonance and baseline models to the low-rank representation and created the fitted spectrum ($s' \in \mathbb{R}^{512}$). The training set consisted of *N*=85,661 frequency domain spectra each consisting of 512. The mean and standard deviation of the amplitude, μ_{train} and σ_{train}, were computed across the entire set, and each input spectrum, $\vec{s}$, was normalized as:

$$\vec{s}_{norm} = \frac{\vec{s} - \mu_{train}}{4 * \sigma_{train}} \tag{10}$$

While CEMD followed an encode-decode scheme, it consisted of two serial encoder-decoder stages. First, the normalized spectrum, $\vec{s}_{norm}$, was passed through a CNN that computed the 32 coarse wavelet coefficients, θ_B. This CNN consisted of 13 convolution

layers, with max-pooling after the 4th, 8th, and 13th layers, followed by two fully-connected (FC) layers (123,128). An estimate of the baseline, $\vec{s}_{baseline}$, was made from the decoder using the wavelet reconstruction technique described in Equations 5-9. Next, the baseline was subtracted from the input spectrum:

$$\vec{s}_{sub} = \vec{s}_{norm} - \vec{s}_{baseline} \tag{11}$$

The baseline-subtracted spectrum, $\vec{s}_{sub}$, was passed through a second CNN consisting of just 6 fully-connected layers that computed the metabolite resonance peak parameters, θ_P. An estimate of the metabolite resonances, $\vec{s}_{peak}$, was made by the decoder using Equation 4. Next, the baseline and resonance peak estimates were added together to produce an estimate of the fitted spectrum, $\vec{s}_{fit}$. The root-mean-squared error (RMSE) of the input and fitted spectrum was used to update the weights of the encoder CNN, and was calculated as:

$$s_{resid} = \sum_{i=1}^{512} \left| s_{norm,i} - s_{fit,i} \right| \tag{12}$$

CEMD was developed in the Python programming language using the TensorFlow 1.3 (Google LLC, Mountain View, CA) library, and trained using TensorFlow's Adam optimizer (78) on a high-end workstation with two Titan X graphical processing units (GPUs; Nvidia Corporation, Santa Clara, CA).

In each epoch of training, spectra in the training data set were run through CEMD to produce fitted spectra, RMSE for each spectrum was calculated, and gradient backpropagation was performed to update the encoder weights. Then, the spectra in the

validation set were run through the CEMD and the validation loss was calculated as the sum of RMSEs. Training continued through multiple epochs until the validation loss converged. Once the autoencoder was trained and the CNN weights finalized, the testing set was used to determine final statistics of CEMD performance. The RMSE was calculated for each spectrum in the testing set, and the mean and standard deviation were reported. Once training of the CEMD encoder weights was complete, the encoder can be applied to spectra to compute the relative concentration of each metabolite resonance based on the parameters in the encoder output, θ_P.

1.24.3 Whole-brain Mapping

A software pipeline to perform CEMD fitting on whole-brain MRSI and to generate volumetric metabolite and ratios maps was developed. Only voxels within the region defined from the brain mask were analyzed, for both the CEMD and MIDAS fitting. While training of the CNN required a GPU, the final CEMD was implemented on a central processing unit (CPU) architecture consisting of a four-core CPU. To assess the utility of generated volumetric for radiation treatment planning, the Cho/NAA ratio map was computed for 10 subjects with glioblastoma fitted by CEMD and an existing parametric analysis method implemented in MIDAS. Spectral fitting in MIDAS used the METAFIT option, which applies three applications of fitting (the FITT program). First, B0 and phase corrections are performed, prior to applying a spatial smoothing and fitting of a higher-SNR copy of the data. Using the initial values from this intermediate result a final spectral analysis is performed on the original spectra. After fitting, voxels were excluded from both

sets of results based on spectral outlier filters, namely those having values that are more than four standard deviations from the mean value within the brain.

Identification of abnormal Cho/NAA regions was determined from the results of each fitting method using the largest single connected component of voxels that had a Cho/NAA at least twofold increased compared to the mean value in contralateral normal-appearing white matter; this particular threshold was previously determined to be optimal for identifying high-risk regions for disease recurrence (20,100,113). The identified regions from each result were compared using the Dice similarity coefficient (DSC) (129) and a Z test was performed for each subject using the logit transform of the DSC (130).

1.25 Results

Training time for the CEMD was approximately 4 hours using TensorFlow on a workstation with two Nvidia Titan X GPUs. The testing set achieved a mean RMSE of fit of 5.0% normalized to the amplitude of the largest peak in each spectrum in the testing set, with a standard deviation of 0.6%. Sample spectra from the testing set are shown in Figure 0.4 with the baseline (red) and peak + baseline (black) fit overlaid on the input spectra (gray). CEMD can handle a variety of baseline effects, varying from a relatively flat baseline near the peaks of interest (Figure 0.4a) to major shifts that can occur at frequencies in the region of lipid or metabolites (Figure 0.4b). Phases can also be determined by the model (Figure 0.4c,d**)**. Even if the signal-to-noise ratio is poor, due to partial volume effects, magnetic field inhomogeneity, or receiver coil sensitivity, CEMD can identify and fit the peak and baseline components.

Sample spectra from three subjects with glioblastoma, not included in the training, testing, or validation sets, are shown in Figure 0.5. In patients with glioblastoma, voxels within the region of active tumor exhibit an increase in Cho and a concomitant decrease in NAA (20). The CEMD-fitted spectra (black) are overlaid on the input spectra (gray). Subject one (Figure 0.5a,b) was scanned at Emory University; subject two (Figure 0.5c,d) was scanned at the University of Miami; and subject three (Figure 0.5e,f) was scanned at the Johns Hopkins University. All three subjects were scanned using the protocol defined in the Methods section, and data were preprocessed in MIDAS as described. Spectra in the left column (Figure 0.5a,c,e) are from voxels in the contralateral normal-appearing hemisphere, while spectra in the right column (Figure 0.5b,d,f) are taken from voxels in regions of tumor.

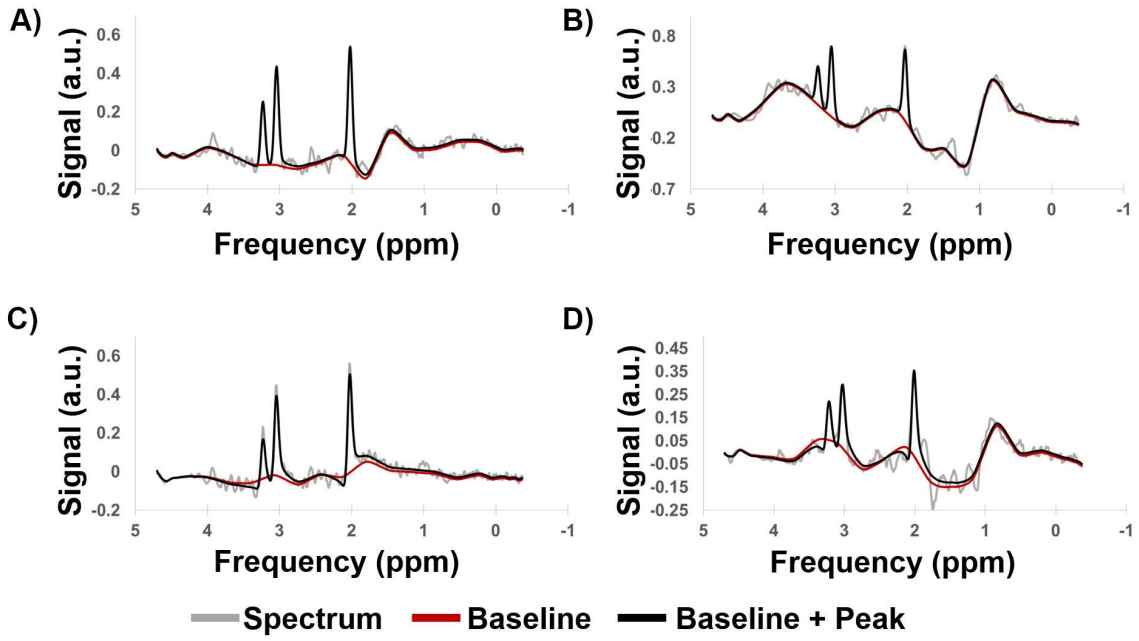


Figure 0.4: Example spectral fittings generated by CEMD. The real components of the computed baseline (red) and baseline + peak (black) fits are overlaid on spectra (gray). Four different types of baseline and phase shifts are shown to indicate that CEMD is able to handle a range of input spectra. a.u. = arbitrary units.

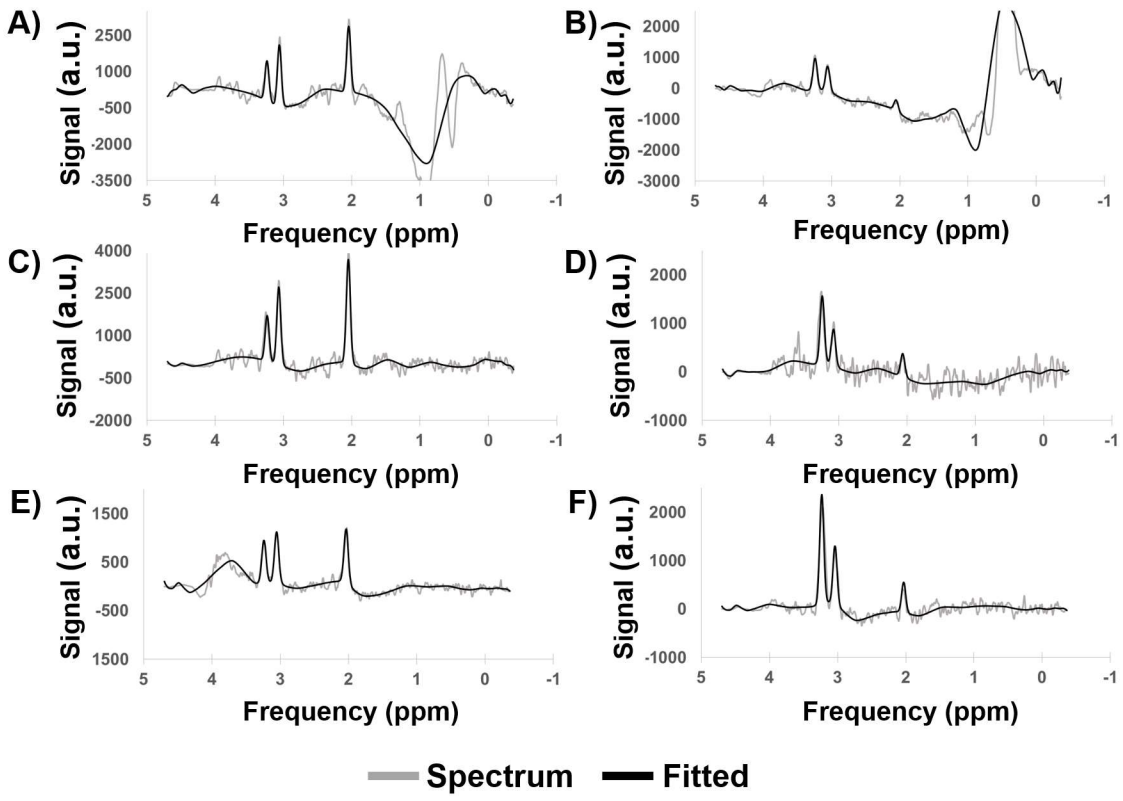


Figure 0.5: Sample spectra (real components) from scans of subjects with glioblastoma. Subject 1 (A,B) was scanned at Emory University; subject 2 (C,D) was scanned at the University of Miami; and subject 3 (E,F) was scanned at Johns Hopkins University. Spectra from regions of healthy tissue (A,C,E) and tumor (B,D,F) are shown. a.u. = arbitrary units.

Correlations between the metabolite concentrations and the Cho/NAA ratio, calculated by CEMD and MIDAS for the testing set are shown in Figure 0.6. The solid blue lines plot the mean value between the two fitting techniques for each bin of values, and the shaded blue region indicates +/- 1 standard deviation. Overlaid in light gray are the histograms for the distribution of metabolite values computed by MIDAS. The variance of CEMD predictions compared to MIDAS is inversely correlated with the number of training samples available. While the two algorithms are linear in the regions where most of the data are present, CEMD has greater uncertainty at the tail end of the histograms where there were fewer training data.

Figure 0.7 compares the fits by MIDAS (blue) and CEMD (black) for several challenging spectra from subjects with glioblastoma. The first (Figure 0.7a) depicts a spectrum with a large baseline shift on the right-side of the NAA peak. The second (Figure 0.7b) shows a spectrum with a large decline of the baseline on the Cho and Cr peaks but a flat baseline near the NAA peak. The third **(**Figure 0.7c) shows a spectrum from a voxel near the inferior rim of the surgical cavity, where partial volume effects reduce the apparent signal-to-noise ratio and where there is an absence of the NAA peak at 2.0 ppm. The fourth (Figure 0.7d) depicts a spectrum from the bone-cortex interface in the temporal; all peaks have broadened linewidth and the NAA peak is adjacent to large noise and lipid signal. All results show comparable performance for fitting of the metabolite peaks with some differences in the baselines.

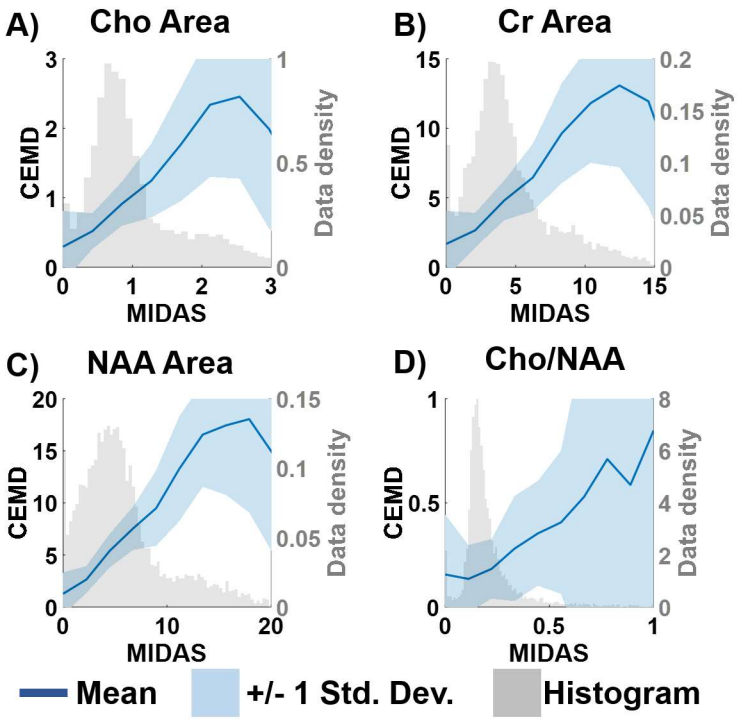


Figure 0.6: Comparison of metabolite values computed by MIDAS and CEMD. The gray histogram shows the distribution of MIDAS calculations; the dark blue line indicates the mean values of CEMD and MIDAS in each histogram bin; the shaded blue region indicates the +/- 1 standard deviation of CEMD calculations.

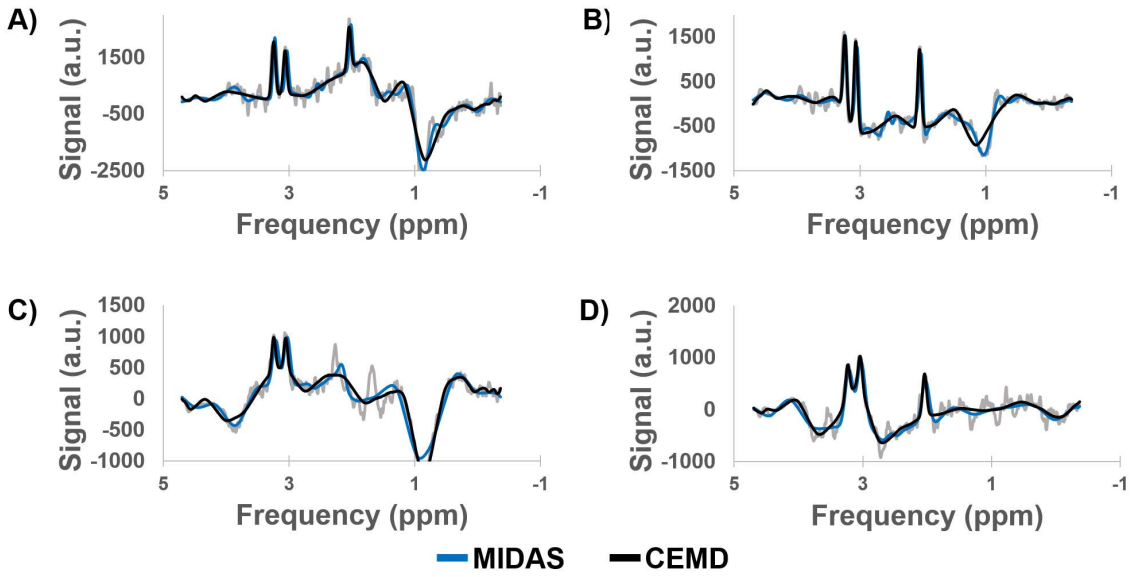


Figure 0.7: Comparison of CEMD and MIDAS spectral fittings on challenging spectra. Spectra were taken from patients with glioblastoma. a.u. = arbitrary units.

Results of the CEMD analysis for studies of a subject with glioblastoma, not included in the training set, are shown in Figure 0.8**,** which shows the individual metabolite maps, the Cho/NAA ratio map, and corresponding contrast-enhanced T1-weighted (CE-T1w) and fluid-attenuated inversion recovery (FLAIR) MRI volumes. Superimposed on the CE-T1w image is a contour drawn by a neuroradiologist to indicate contrast enhancing tissue and the surgical cavity, regions that would normally be targeted for high dose radiation therapy.

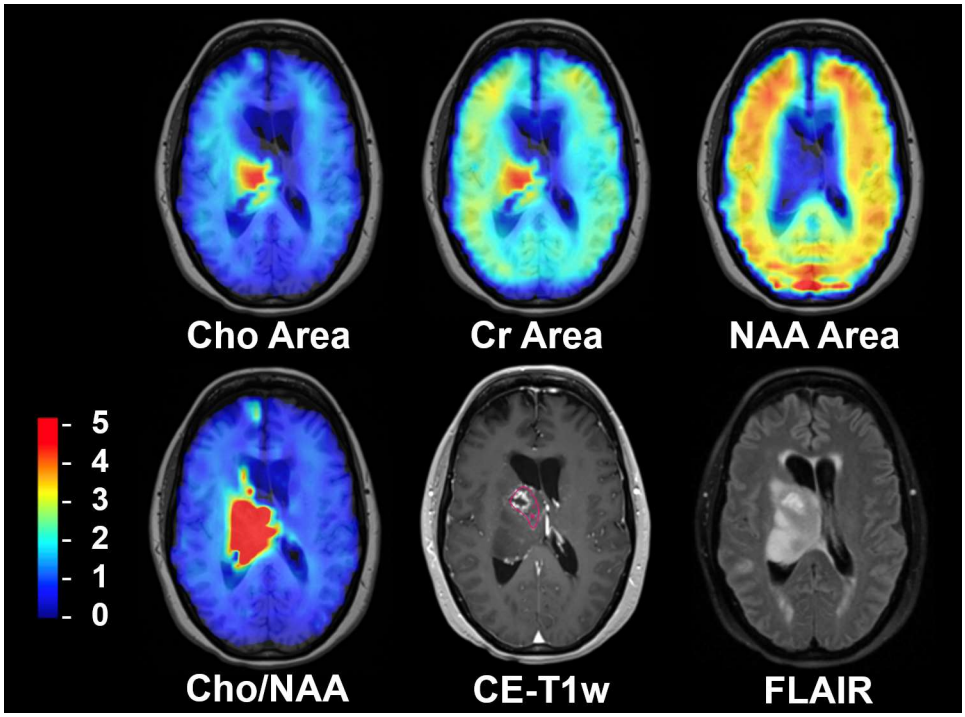


Figure 0.8: Example whole-brain metabolite maps generated by CEMD for a patient with glioblastoma. The Cho/NAA volume indicates the presence of metabolically active tumor around the resection cavity extending beyond the contrast-enhancing lesion. Color bar corresponding to relative Cho/NAA values (compared to contralateral normal-appearing white matter) for Cho/NAA image; arbitrary units for metabolite maps. Pink contour indicated neuroradiologist-segmented contrast-enhancing tissue around the resection cavity.

Cho/NAA abnormality volumes were contoured for CEMD and MIDAS fitting and the results are shown in Table 0.1, which includes subject-wise execution time, abnormality volumes, and DSC. The mean execution time for whole-brain spectral fitting was 20.6 +/- 2.8 sec using the CEMD. As a representative example, the execution time for CEMD on Subject 6 was broken down as follows: 2.7 seconds to load 11,702 spectra from disk into memory; 17.0 seconds to load the CEMD encoder model and to process all spectra; 3.3 seconds to create volumetric maps for each metabolite and the Cho/NAA ratio and to write these maps to disk. On average, CEMD-derived lesion volumes are larger by 1.2 cm^3 (P = 0.83 using a two-tailed paired T test) and in several subjects encapsulated the contours produced from the MIDAS analysis. The mean DSC between CEMD and MIDAS was 0.72 +/- 0.13. In Figure 0.9 and Figure 0.10, Cho/NAA volumes computed by MIDAS and CEMD for Subjects 9 and 1, respectively, were qualitatively compared; the contours indicate a twofold increase in Cho/NAA compared to contralateral white matter. Isolated bright spots are artifacts due to fitting errors and were not contoured. A spectrum from an area where the two algorithms had different values of Cho/NAA (white box) is shown for each subject, with CEMD fit (black) and MIDAS fit (blue) overlaid on the spectrum (gray). Discrepancies occur either due to the calculated ratio being just above or below the 2x threshold (Figure 0.10b), or in areas of minimal or poor signal quality and therefore the uncertainties in the measurements must be considered to be very high.

Table 0.1: Subject-wise comparison of CEMD and MIDAS.

Subject	CEMD Time (sec)	2x Cho/NAA Vol. (cm^3)		Volume Difference (cm^3)	DSC	Z Test P Value
		CEMD	MIDAS			
1	17.8	57.9	56.7	1.2	0.84	0.047
2	26.3	31.9	35.1	-3.2	0.56	0.410
3	18.1	72.2	76.0	-3.8	0.87	0.026
4	21.0	35.9	39.9	-4.1	0.65	0.265
5	17.4	23.0	39.6	-16.7	0.64	0.277
6	23.0	5.9	4.0	1.8	0.59	0.366
7	19.0	105.2	134.1	-28.8	0.86	0.037
8	20.4	86.9	60.0	26.9	0.75	0.141
9	20.1	54.9	22.8	32.1	0.57	0.385
10	23.1	37.7	30.7	7.0	0.81	0.072
Mean	**20.6**	**--**	**--**	**1.2**	**0.72**	**--**
Std. Dev	**2.8**	**--**	**--**	**18.1**	**0.13**	**--**

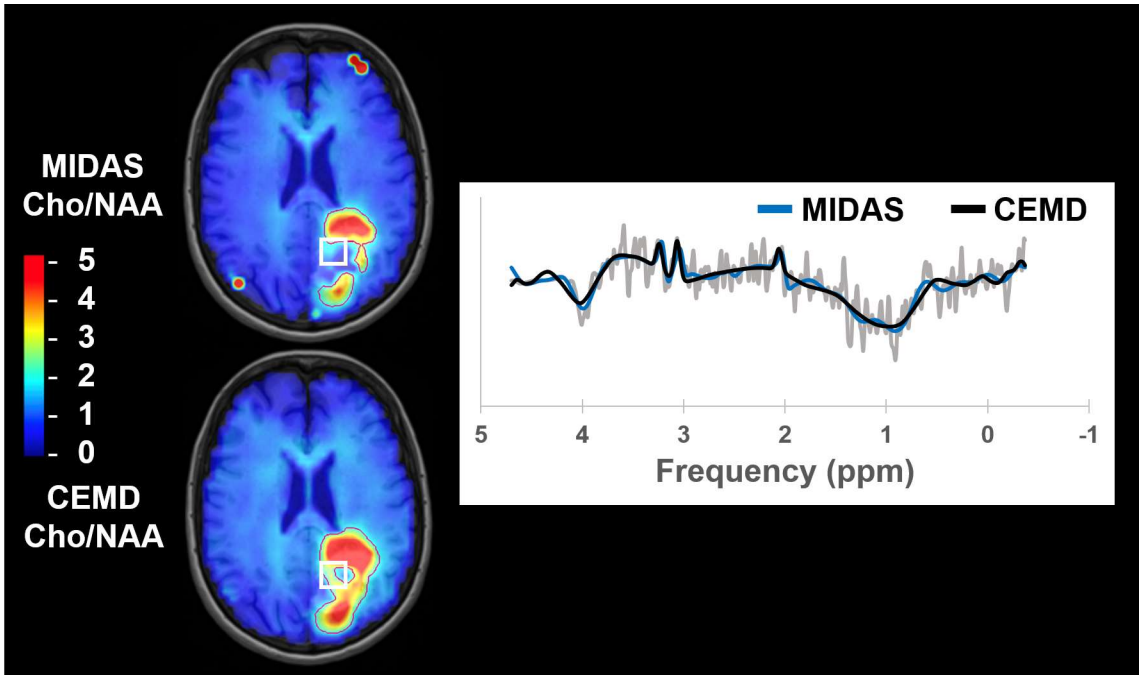


Figure 0.9: Comparison of the Cho/NAA volumes generated by MIDAS and CEMD in a subject with glioblastoma. Contours indicate Cho/NAA values greater than two-fold the value in contralateral normal appearing white matter. The spectrum shown comes from the highlighted voxel (white box).

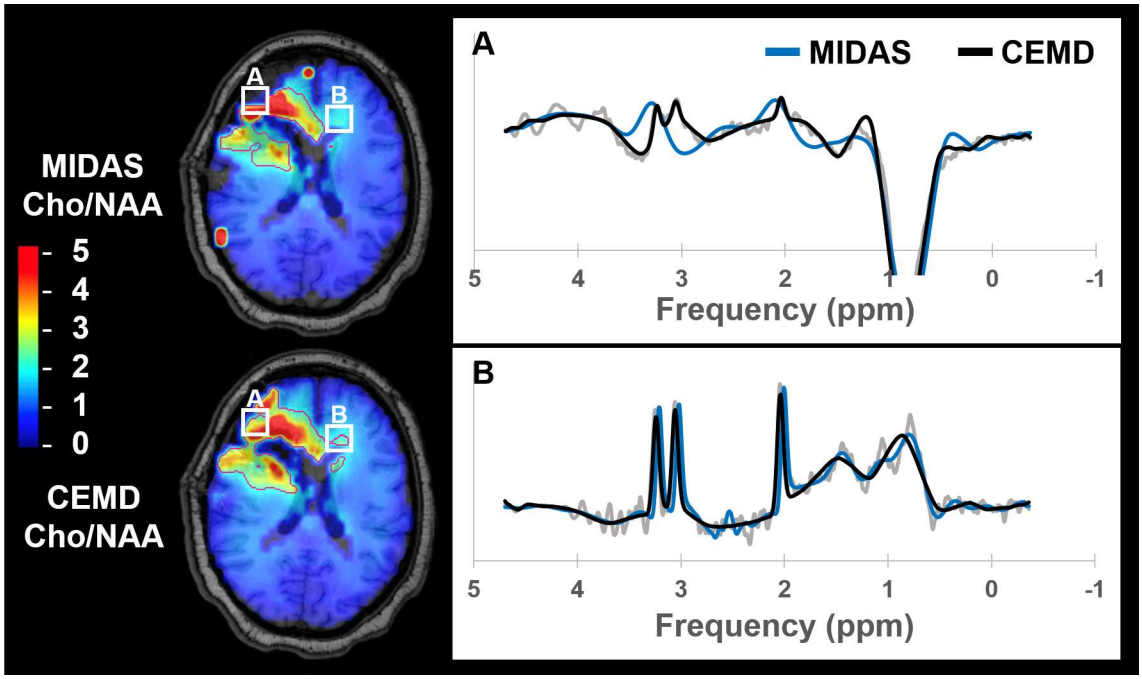


Figure 0.10: A second example of CEMD and MIDAS-generated metabolite maps. A comparison of Cho/NAA volumes generated by MIDAS (blue) and CEMD (black) in a second subject with glioblastoma, with spectra from the highlighted voxels (white boxes) shown. Contours indicate the isoline for Cho/NAA values greater than two-fold the value in contralateral normal appearing white matter.

1.26 Discussion

In recent years, machine learning has seen tremendous advances and has shattered benchmarks in a variety of fields, including medical imaging applications (128). While these deep learning approaches, including CNNs, can outperform less complex models, a key issue is that of interpretability. Several techniques, such as gradient-weighted class activation mapping (81), seek to elucidate some of the internal workings of CNNs and provide insight as to *why* the CNN predicted a particular output. Even so, this is always done in a retrospective fashion after the CNN has already been trained, and thus insights cannot be used to modify the algorithm. It is difficult to incorporate a priori domain knowledge into deep learning because these techniques are fundamentally data-driven rather than model or knowledge driven. However, recent work has suggested that the incorporation of domain knowledge may be able to improve the performance of deep learning models (131,132).

In this work, a deep learning approach to spectral fitting that incorporates a priori spectral information was developed and evaluated. Spectral fitting is the computational bottleneck in processing of volumetric MRSI, largely because the existing methods are based on iterative algorithms. The CEMD is an unsupervised deep learning architecture that incorporates spectral models to generate an encoding of spectral parameters, which is advantageous because it does not require any "ground truth" spectral quantitation for training. The predictions of the CEMD have contextual meaning, and the CEMD was trained to make these predictions within the constraints of an explicitly defined spectral model. Once trained, the CEMD performs spectral fitting on volumetric data in under one minute using standard computer hardware. The order of magnitude improvement in fitting

time can greatly benefit the clinical adoption of whole-brain MRSI. This architecture could be implemented on scanner computers, enabling real-time reconstruction and review of data without the need for offloading of data to more specialized computer systems. Fitted data and metabolite volumes could then easily be sent to a clinical PACS systems in a streamlined fashion.

Incorporation of knowledge of the spectral model imposes constraints that are important for the assessment of spectra. Bhat et. al. previously incorporated spectral models in an unsupervised neural network; however, their model was limited in requiring baseline-corrected spectra as input to their neural network (59). CEMD simultaneously computes the baseline and peak components of the spectrum and correctly identifies singlets in challenging spectra such as those in Figure 0.4. The spectrum in Figure 0.4c indicates the loss of the NAA resonance at 2.0 ppm. It is necessary for a fitting model understand where the NAA resonance should be and not attempt to model other peak-like shapes (e.g. at 2.3 ppm and 1.7 ppm) as NAA. For spectra where asymmetric broadening of peaks occurs, the spectral lineshape model dictates that a symmetric peak should be fit (Figure 0.4d). In these cases, CEMD performed on par with traditional parametric analysis algorithms such as that incorporated in MIDAS.

For this study only the real part of the complex spectral data was used as input for CEMD and in the cost function for updating the CNN weights, though a complex spectral model (Equation 4) was used in the decoder. We note that the phase correction terms were accurately reported. It is speculated that this occurs through an observation from the asymmetry of the lineshape (133). An initial implementation of CEMD using the full

complex spectral data found reduced performance (mean fit RMSE of 38%), potentially due to the increased number of parameters in the encoder that would need to be trained.

While the CEMD was trained using data obtained on a single 3T scanner, this study has demonstrated generalizability to data acquired on other instruments using the same acquisition parameters (Figure 0.5); however, this study has not evaluated the extension to other pulse sequence parameters. If studies are to be performed using different acquisition schemes, such as with shorter TE, CEMD would have to be retrained. However, training does not require many subjects because of the ~10,000 spectra in each study. The CEMD could be adapted, e.g. changing the number and type of layers in the encoder and number of coefficients in the decoder models, for more resonances, of multiplets of resonances, or of metabolites whose resonance peaks are not readily separable (gluatamate and glutamine). In general, because the autoencoder scheme does not rely on any external "ground truth" data, it can be readily adapted and optimized for different complexities of fitting.

In this work, a comparison between this new fitting paradigm and an existing iterative parametric optimization method was performed. The results in Figure 0.6 show the correlation between the two algorithms' estimation of Cho, Cr, and NAA resonances and the Cho/NAA ratio on the same set of spectra. In spectroscopy, the acquired spectral signal amplitude is uncalibrated, and additional methods are required to apply a signal normalization procedure. Ratios, such as Cho/NAA, do not require calibration; however, they are more unstable when the denominator is low. As seen in Figure 0.7d, both CEMD and MIDAS fitting have high uncertainty for high Cho/NAA values that represent low NAA regions. While individual metabolite resonance maps are shown in Figure 0.8 to

assess the ability of CEMD to perform whole-brain fitting, the Cho/NAA maps are used as the main comparison between CEMD and an existing parametric analysis (MIDAS) in Table 0.1. CEMD achieves a Dice coefficient of 0.72 when compared to MIDAS. A key observation is that CEMD can produce similar Cho/NAA volumes to MIDAS while never being trained to do so; thus, it independently achieves similar results, which highlights the power of unsupervised learning techniques.

Qualitatively, as seen in Figure 0.9 and Figure 0.10, CEMD produces similar spatial distributions of the Cho/NAA ratio as MIDAS in subjects with glioblastoma. Both algorithms identify the region of brain with an elevated ratio, though occasionally with different contour shape and size when selecting voxels with a Cho/NAA abnormality greater than 2x contralateral NAWM (Table 0.1). These differences hold when adjusting this threshold on two sample cases, one from a subject with high DSC conformality and another with low DSC conformality (Table 0.2). The study for Subject 9 has poor DSC; reviewing spectra in which the two fitting algorithms produced different values of Cho/NAA reveals that these discordant fits occur in regions of low spectral SNR surrounding the rim of the surgical cavity. Two possible reasons underlying this discordance are speculated. The first is highlighted in Figure 0.9, in which CEMD estimates a slightly lower value for the NAA resonance peak compared to MIDAS, and therefore estimates a Cho/NAA above the 2x threshold. A similar discrepancy in the Cho/NAA value can be observed in Subject 1 in the contralateral extension of the CEMD contour in Figure 0.10b. While using a threshold of a twofold increase in Cho/NAA compared to NAWM has been previously established to identify a high probability of tumor, a voxel having a value just below the threshold does not mean the voxel is tumor-

free. Ultimately, when using MRSI for therapeutic guidance or diagnosis, clinician insight is needed.

Table 0.2: Dice coefficients of contours for different Cho/NAA thresholds.

Subject 3			
Rel. Cho/NAA Threshold	**1.75**	**2.0**	**2.25**
DSC	0.86	0.87	0.88
Z test p value	0.04	0.03	0.02
Subject 6			
Rel. Cho/NAA Threshold	**1.75**	**2.0**	**2.25**
DSC	0.55	0.59	0.51
Z test p value	0.43	0.37	0.49

A second reason for discordance between the two algorithms is apparent in a voxel from Subject 1, shown in Figure 0.10a. CEMD has a tendency to overfit spectra with low SNR that occur near tissue interfaces, where magnetic field inhomogeneities and partial volume effects can cause large distortions. The highlighted voxel is deemed to be artifactual and should in fact have been removed. In general, results corresponding to low SNR data, as shown in Figures 7 and 8, can be excluded by post-processing that takes into account the uncertainty of the fit via metrics such as Cramér-Rao lower bounds (26,45). This is especially important when calculating metabolite ratios. One limitation of the current implementation of the CEMD fitting is that it does not include an estimate of the uncertainties of the fit and this needs to be further investigated, although our previously-described artifact filter may also be used for this purpose (119). While CEMD was built

for spectral fitting for rapid turnaround time, for research purposes in which a more thorough analysis of spectral quantitation is desired, existing software such as MIDAS can be used.

1.27 Conclusion

In this work, a machine learning approach to spectral fitting is described that can perform sub-minute calculation of relative metabolite concentrations in MRSI of the brain. A convolutional encoder-model decoder technique has been implemented that explicitly incorporates a standard parametric spectral model with the power of unsupervised feature-learning to produce fast spectral fittings that are constrained by the standard model. This is a powerful paradigm that does not require a priori ground truth and relies upon previously-used spectral lineshape and baseline models to optimize the underlying convolutional neural network parameters. The CEMD architecture is shown to produce accurate fitting of a variety of spectra acquired from multiple scanners in patients with glioblastoma, including correctly fitting challenging spectra with low SNR, partial volume effects, baseline shifts, phase shifts, and dropout of one or more metabolite resonances. The CEMD can fit whole-brain data on a standard multicore computer without the need for expensive workstations or GPUs, in less than one minute. With this new autoencoder-based neural network, the largest computational bottleneck in processing MRSI can be overcome, bringing improved performance that will support the implementation of MRS for more widespread clinical use.

CONCLUSION

Proton spectroscopic magnetic resonance imaging (sMRI) is an advanced imaging modality that enables the non-invasive detection of in vivo tissue metabolites without the need for any exogenous contrast agents. It has been shown to provide diagnostic and prognostic value for several neuropathologies, particularly highly proliferative disease such as glioblastoma. While it has been a growing area of research, it's implementation clinically has been limited due to several technical challenges. In this dissertation, we sought to provide engineering and informatics solutions to four of these challenges.

In Chapter 2, we present our work on developing a machine learning approach for filtering spectral artifacts in whole-brain sMRI. Artifacts are spectra which are of poor quality such that accurate quantitation of metabolites cannot be performed. Artifacts are due to a variety of causes, including partial volumes, transitions between types at tissue boundaries, local magnetic field inhomogeneities, subject motion, low signal-to-noise ratio (SNR) causing peaks to be washed out, and improper water or lipid suppression (43). When viewing spatial heat maps for sMRI, as shown in Figure 0.2, artifacts can be appear as hypo- or hyperintense foci. In order to use sMRI for diagnosis or treatment planning, it is important to ensure that erroneous spectra are removed prior to evaluation. The gold standard is manual review by an MR spectroscopy expert, and while this is sufficient for single and multi-voxel MRS, it does not scale well to whole-brain imaging which contain on the order of 10,000 independent spectra.

Several statistical techniques have been developed to identify artifactual spectra, such as measuring linewidths, Cramer-Rao bounds of spectral fitting, SNR analysis, magnitude

of peaks, measurement of tissue fraction within a voxel, and detection of insufficient water or lipid suppression (18,26,45). A combination of these metrics is, for example, implemented in the MIDAS software package, a commonly used application for whole-brain sMRI processing. However, a percentage of voxels will still make it through these filters and will require manual review.

To address this issue, we developed a convolutional neural network (CNN) architecture that sought to mimic the implicit evaluation criteria used by MRS experts. While most experts will use some baseline criteria for evaluating spectral quality (e.g. SNR or linewidth), there is also some higher-level features they use that cannot be explicitly modeled but is rather learned from experience. We designed a data collection program (OSCAR) in a manner that enabled a consensus classification of spectra from a panel of experts. These data were then used to train a CNN to classify spectra as being of "good" or "poor" quality. The CNN showed high sensitivity and specificity (AUC=0.95) and showed no preference towards any single expert. We have since implemented the CNN into our workflows and can use it when needed to filter out artifactual voxels.

There are several improvements that can be made to this CNN in future work. When tuned to the optimum sensitivity/specificity threshold, the CNN is potentially too aggressive in removing voxels near tumor necrotic cavities, as seen in Figure 0.8. During data collection with OSCAR, the expert raters were not given any indication of pathology beyond the location of the voxel on a T1w slice, and no data regarding pathology was explicitly fed to the CNN. An improved architecture may include, for example, a binary segmentation of tumor vs non-tumor during data collection and as part of the input to the CNN. The spatial distribution of a given metabolite typically does not produce isolated

single-voxel differences (it is smooth), whereas an artifact may be limited to a single voxel. Taking into account information about neighboring voxels, therefore, could be of use in improving detection of artifacts, particularly in and around tumor. Finally, in our work we took the approach of using as input non-fitted spectra, such that the algorithm could be applied to any pipeline. However, as noted by the success of several other groups (54,55,66), features generated by (and dependent on) spectral fitting do contain information that is useful for classifying spectral quality. A future model could combine the unbiased deep learning approach used in our work with some engineered features from spectral fitting.

In Chapter 3, we develop an algorithm to internally normalize a subject's sMRI study by contralateral normal-appearing white matter. This technique uses the Cho/NAA map, which has a two-Gaussian distribution, to separate out healthy vs pathologic voxels, and then calculates the mean of the healthy voxels in contralateral white matter. A segmentation of the selected voxels can also be applied to other metabolites to generate NAWM-normalized maps for each of them. This approach has several advantages in that it is automated, fast, and not-subjective. However, it does require a high-coverage Cho/NAA map, which may not be present in all subjects, and it does not work in cases where there is only a single Gaussian source present (e.g. low grade gliomas, healthy controls, other non-proliferative pathology). To address these, an algorithm that does not depend on any particular metabolite would need to be developed. In the time since the work in Chapter 3 was completed, the developers of the MIDAS software have added such an algorithm, which uses tissue fractions of gray matter to detect tumor, since tumor tends to destroy

gray matter as visible on T1w MRI, and then select white matter voxels in the opposite hemisphere.

Chapter 3 also introduced BrICS, the Brain Imaging Collaboration Suite. This browser-based application has enabled clinicians from Emory and other institutions to visualize, annotate, and evaluate sMRI for subjects with several pathologies. BrICS is the primary software platform being used for NCT03137888, and using it, we have been able to show the feasibility of incorporated spectroscopic information when planning radiation therapy targets. At the time of writing this Chapter, enrollment is complete for this study, and an analysis will be presented after sufficient follow-up data are collected for each subject.

Since the publication of the work in Chapter 3, we have continued development of BrICS to be a platform for more neuropathologies. There are new algorithms for segmentation based on T1w, FLAIR, and sMRI information, and the software is able to handle any metabolite or combination maps (e.g. ratio or multiplication). We have also developed an integration with TensorFlow, such that our existing and future machine learning algorithms can be easily called from BrICS. Finally, we have built an integration with the REDCap database software package, enabling querying of de-identified data from within the HIPAA-compliant database package. BrICS is now deployable as a Docker container (source code available upon request at **https://github.com/sgurbani/brics**), and is being developed to be deployable at scale for larger clinical studies. Once the outcomes of NCT03137888 have been analyzed, we plan to submit a national consortium trial to assess the efficacy of sMRI-guided dose escalation radiation therapy for patients with glioblastoma.

In Chapter 4, we introduced a novel unsupervised architecture to perform spectral quantitation, the convolutional encoder-model decoder (CEMD). In effect, CEMD mimics the expectation-maximization task that is performed with iterative algorithms. The advantages of this technique over prior work are: 1) it is unsupervised, and so does not need to rely on "correct" (e.g. manually reviewed) spectral quantitations; 2) it is constrained within known models of magnetic resonance, thereby enabling ready interpretability of results; 3) it can correctly model slight frequency shifts from known resonances and correctly account for phase shifts; and 4) it can correctly model spectra with a variety of baseline signals, including insufficient water and lipid suppression. Once CEMD is trained for a particular set of acquisition parameters, it can perform whole-brain fitting in less than 30 seconds, a full two orders of magnitude improvement over the current state-of-the-art. In the future, CEMD can be expanded to incorporate a broader set of metabolites, multiplets of metabolites, and to use different baseline functions if desired. It is our goal to one day have CEMD implemented directly on MR instrumentation, such that the results of an sMRI scan can be viewed immediately after acquisition, as is the case for other MR sequences.

With these four improvements, we have addressed several of the technical challenges that have hindered the clinical implementation of sMRI. We have shown the feasibility of using sMRI in the clinical research setting via NCT03137888, and we are preparing for a much larger consortium trial to start in the next few years. With continued technical development, perhaps sMRI may become a part of the standard-of-care imaging that patients with glioblastoma will undergo prior to their treatment.

REFERENCES

1. Gunderman RB, Siddiqui AR, Heitkamp DE, Kipfer HD. The Vital Role of Radiology in the Medical School Curriculum. American Journal of Roentgenology 2003;180(5):1239-1242.

2. Jäkel S, Dimou L. Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation. Frontiers in cellular neuroscience 2017;11:24-24.

3. Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta neuropathologica 2016;131(6):803-820.

4. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. Neuro-Oncology 2017;19(suppl_5):v1-v88.

5. Stupp R, Hegi ME, Mason WP, van den Bent MJ. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. The lancet oncology 2009.

6. Zlokovic BV. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. Nature reviews Neuroscience 2011;12(12):723-738.

7. McRobbie DW, Moore EA, Graves MJ, Prince MR. MRI from Picture to Proton: Cambridge university press; 2017.

8. Bellin M-F, Van Der Molen AJ. Extracellular gadolinium-based contrast media: an overview. European journal of radiology 2008;66(2):160-167.

9. Wernicke AG, Smith AW, Taube S, Mehta MP. Glioblastoma: Radiation treatment margins, how small is large enough? Practical radiation oncology 2016;6(5):298-305.

10. Tsuchiya K, Mizutani Y, Hachiya J. Preliminary evaluation of fluid-attenuated inversion-recovery MR in the diagnosis of intracranial tumors. American journal of neuroradiology 1996;17(6):1081-1086.

11. Weller M, Cloughesy T, Perry JR, Wick W. Standards of care for treatment of recurrent glioblastoma—are we there yet? Neuro-Oncology 2013;15(1):4-27.

12. Campos B, Olsen LR, Urup T, Poulsen HS. A comprehensive profile of recurrent glioblastoma. Oncogene 2016;35:5819.

13. Miller BL. A review of chemical issues in 1H NMR spectroscopy: N-acetyl-l-aspartate, creatine and choline. NMR in Biomedicine 1991.

14. He Q, Xu RZ, Shkarin P, Pizzorno G, Lee-French CH, Rothman DL, Shungu DC, Shim H. Magnetic resonance spectroscopic imaging of tumor metabolic markers for cancer diagnosis, metabolic phenotyping, and characterization of tumor microenvironment. Disease markers 2004;19(2-3):69-94.

15. Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo 2000;13(3):129-153.

16. Mansfield P, Maudsley AA, Bains T. Fast scan proton density imaging by NMR. Journal of Physics E: Scientific Instruments 1976;9(4):271.

17. Maudsley AA, Domenig C, Govind V, Darkazanli A, Studholme C, Arheart K, Bloomer C. Mapping of brain metabolite distributions by volumetric proton MR spectroscopic imaging (MRSI). Magn Reson Med 2009;61(3):548-559.

18. Sabati M, Sheriff S, Gu M, Wei J, Zhu H, Barker PB, Spielman DM, Alger JR, Maudsley AA. Multivendor implementation and comparison of volumetric whole-brain echo-planar MR spectroscopic imaging. Magn Reson Med 2015;74(5):1209-1220.

19. Cordova JS, Gurbani SS, Olson JJ, Liang Z, Cooper LAD, Shu H-KG, Schreibmann E, Neill SG, Hadjipanayis CG, Holder CA, Shim H. A systematic pipeline for the objective comparison of whole-brain spectroscopic MRI with histology in biopsy specimens from grade III glioma. Tomography 2016;2(2):106-116.

20. Cordova JS, Shu H-KG, Liang Z, Gurbani SS, Cooper LAD, Holder CA, Olson JJ, Kairdolf B, Schreibmann E, Neill SG, Hadjipanayis CG, Shim H. Whole-brain spectroscopic MRI biomarkers identify infiltrating margins in glioblastoma patients. Neuro Oncol 2016;18(8):1180-1189.

21. Law M. MR spectroscopy of brain tumors. Topics in Magnetic Resonance Imaging 2004;15(5):291-313.

22. Soares DP, Law M. Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications. Clin Radiol 2009;64(1):12-21.

23. Young K, Govindaraju V, Soher BJ, Maudsley AA. Automated spectral analysis I: formation of a priori information by spectral simulation. Magn Reson Med 1998;40(6):812-815.

24. Armstrong BH. Spectrum line profiles: the Voigt function. Journal of Quantitative Spectroscopy and Radiative Transfer 1967;7(1):61-88.

25. Vanhamme L, van den Boogaart A, Van Huffel S. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. Journal of magnetic resonance 1997;129(1):35-43.

26. Provencher SW. Automatic quantitation of localized in vivo1H spectra with LCModel. NMR Biomed 2001;14(4):260-264.

27. Ratiney H, Sdika M, Coenradie Y, Cavassila S, Ormondt Dv, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo 2005;18(1):1-13.

28. Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. Magnetic Resonance in Medicine 2010;65(1):1-12.

29. Young K, Soher BJ, Maudsley AA. Automated spectral analysis II: application of wavelet shrinkage for characterization of non-parameterized signals. Magn Reson Med 1998;40(6):816-821.

30. Soher BJ, Young K, Govindaraju V, Maudsley AA. Automated spectral analysis III: application to in vivo proton MR spectroscopy and spectroscopic imaging. Magn Reson Med 1998;40(6):822-831.

31. Lam F, Liang ZP. A subspace approach to high-resolution spectroscopic imaging. Magnetic resonance in medicine 2014;71(4):1349-1357.

32. Michaelis T, Merboldt KD, Bruhn H, Hänicke W, Frahm J. Absolute concentrations of metabolites in the adult human brain in vivo: quantification of localized proton MR spectra. Radiology 1993;187(1):219-227.

33. Husted CA, Duijn JH, Matson GB, Maudsley AA, Weiner MW. Molar quantitation of in vivo proton metabolites in human brain with 3D magnetic resonance spectroscopic imaging. Magnetic resonance imaging 1994;12(4):661-667.

34. Christiansen P, Henriksen O, Stubgaard M, Gideon P, Larsson HBW. In vivo quantification of brain metabolites by 1H-MRS using water as an internal standard. Magnetic resonance imaging 1993;11(1):107-118.

35. Lecocq A, Le Fur Y, Maudsley AA, Le Troter A, Sheriff S, Sabati M, Donnadieu M, Confort-Gouny S, Cozzone PJ, Guye M. Whole-brain quantitative mapping of metabolites using short echo three-dimensional proton MRSI. Journal of Magnetic Resonance Imaging 2015;42(2):280-289.

36. van der Knaap MS, van der Grond J, van Rijen PC, Faber JA, Valk J, Willemse K. Age-dependent changes in localized proton and phosphorus MR spectroscopy of the brain. Radiology 1990;176(2):509-515.

37. Angelie E, Bonmartin A, Boudraa A, Gonnaud P-M, Mallet J-J, Sappey-Marinier D. Regional differences and metabolic changes in normal aging of the human brain: proton MR spectroscopic imaging study. American Journal of Neuroradiology 2001;22(1):119-127.

38. Mueller C, Lin JC, Sheriff S, Maudsley AA, Younger JW. Evidence of widespread metabolite abnormalities in Myalgic encephalomyelitis/chronic fatigue syndrome: assessment with whole-brain magnetic resonance spectroscopy. Brain imaging and behavior 2019:1-11.

39. Busch M, Liebenrodt K, Gottfried S, Weiland E, Vollmann W, Mateiescu S, Winter S, Lange S, Sahinbas H, Baier J. Influence of brain tumors on the MR spectra of healthy brain tissue. Magnetic resonance in medicine 2011;65(1):18-27.

40. Goebell E, Fiehler J, Ding XQ, Paustenbach S, Nietz S, Heese O, Kucinski T, Hagel C, Westphal M, Zeumer H. Disarrangement of fiber tracts and decline of neuronal density correlate in glioma patients—a combined diffusion tensor imaging and 1H-MR spectroscopy study. American journal of neuroradiology 2006;27(7):1426-1431.

41. Maudsley AA, Roy B, Gupta RK, Sheriff S, Awasthi R, Gu M, Husain N, Mohakud S, Behari S, Spielman DM. Association of metabolite concentrations and water diffusivity in normal appearing brain tissue with glioma grade. Journal of Neuroimaging 2014;24(6):585-589.

42. McKnight TR, von dem Bussche MH, Vigneron DB, Lu Y, Berger MS, McDermott MW, Dillon WP, Graves EE, Pirzkall A, Nelson SJ. Histopathological validation of a three-dimensional magnetic resonance spectroscopy index as a predictor of tumor presence. Journal of Neurosurgery 2002;97(4):794-802.

43. Kreis R. Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts. NMR Biomed 2004;17(6):361-381.

44. Menze BH, Kelm BM, Weber MA, Bachert P, Hamprecht FA. Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images. Magn Reson Med 2008;59(6):1457-1466.

45. Jiru F, Skoch A, Klose U, Grodd W, Hajek M. Error images for spectroscopic imaging by LCModel using Cramer-Rao bounds. MAGMA 2006;19(1):1-14.

46. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer series in statistics New York; 2001.

47. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods 2000;43(1):3-31.

48. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. 2012. p 1097-1105.

49. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436-444.

50. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. Human-level control through deep reinforcement learning. Nature 2015;518(7540):529-533.

51. Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans Image Process 2012;21(3):968-982.

52. Liao S, Gao Y, Oto A, Shen D. Representation learning: a unified deep learning framework for automatic prostate MR segmentation. Med Image Comput Comput Assist Interv 2013;16(Pt 2):254-261.

53. Wright AJ, Arús C, Wijnen JP, Moreno-Torres A, Griffiths JR, Celda B, Howe FA. Automated quality control protocol for MR spectra of brain tumors. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 2008;59(6):1274-1281.

54. Pedrosa de Barros N, McKinley R, Knecht U, Wiest R, Slotboom J. Automatic quality control in clinical (1)H MRSI of brain cancer. NMR Biomed 2016;29(5):563-575.

55. Pedrosa de Barros N, McKinley R, Wiest R, Slotboom J. Improving labeling efficiency in automatic quality control of MRSI data. Magn Reson Med 2017.

56. Kyathanahally SP, Doering A, Kreis R. Ghostbusters for MRS: Automatic Detection of Ghosting Artifacts using Deep Learning. 2017; Honolulu, HI. p Abstract #5479.

57. Kyathanahally SP, Döring A, Kreis R. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. Magnetic resonance in medicine 2018;80(3):851-863.

58. Hiltunen Y, Kaartinen J, Pulkkinen J, Häkkinen A-M, Lundbom N, Kauppinen RA. Quantification of Human Brain Metabolites from in Vivo1H NMR Magnitude Spectra Using Automated Artificial Neural Network Analysis. Journal of Magnetic Resonance 2002;154(1):1-5.

59. Bhat H, Sajja BR, Narayana PA. Fast quantification of proton magnetic resonance spectroscopic imaging with artificial neural networks. Journal of Magnetic Resonance 2006;183(1):110-122.

60. Das D, Coello E, Schulte RF, Menze BH. Quantification of Metabolites in Magnetic Resonance Spectroscopic Imaging Using Machine Learning. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors2017 2017//; Cham. Springer International Publishing. p 462-470.

61. Das D, Coello E, Sekuboyina A, Schulte RF, Menze BH. Direct Estimation of Model Parameters in MR Spectroscopic Imaging using Deep Neural Networks. 2018; Paris, FR. p Abstract #3852.

62. Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology 2002;222(3):715-721.

63. Li X, Jin H, Lu Y, Oh J, Chang S, Nelson SJ. Identification of MRI and 1H MRSI parameters that may predict survival for patients with malignant gliomas. NMR Biomed 2004;17(1):10-20.

64. Kreis R. Issues of spectral quality in clinical 1H magnetic resonance spectroscopy and a gallery of artifacts. NMR Biomed 2004;17(6):361-381.

65. Slotboom J, Nirkko A, Brekenfeld C, Ormondt Dv. Reliability testing of in vivo magnetic resonance spectroscopy (MRS) signals and signal artifact reduction by order statistic filtering. Meas Sci Technol 2009;20(10):104030.

66. Kyathanahally SP, Mocioiu V, Pedrosa de Barros N, Slotboom J, Wright AJ, Julià-Sapé M, Arus C, Kreis R. Quality of clinical brain tumor MR spectra judged by humans and machine learning tools. Magn Reson Med 2017.

67. Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans Image Process 2012;21(3):968-982.

68. Dawes TJW, de Marvao A, Shi W, Fletcher T, Watson GMJ, Wharton J, Rhodes CJ, Howard LSGE, Gibbs JSR, Rueckert D, Cook SA, Wilkins MR, O'Regan DP. Machine Learning of Three-dimensional Right Ventricular Motion Enables Outcome Prediction in Pulmonary Hypertension: A Cardiac MR Imaging Study. Radiology 2017;283(2):381-390.

69. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology 2017;284(2):574-582.

70. Liao S, Gao Y, Oto A, Shen D. Representation learning: a unified deep learning framework for automatic prostate MR segmentation. Med Image Comput Comput Assist Interv 2013;16(Pt 2):254-261.

71. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, Gutman DA, Halani SH, Velazquez Vega JE, Brat DJ, Cooper LAD. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Scientific Reports 2017;7(1):11707.

72. Sainath TN, Weiss RJ, Senior A, Wilson KW, Vinyals O. Learning the speech front-end with raw waveform CLDNNs. Sixteenth Annual Conference of the International Speech Communication Association: ee.columbia.edu; 2015.

73. Wei Dai CDSQJLSD. Very Deep Convolutional Neural Networks for Raw Waveforms. 2016.

74. Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25: Curran Associates, Inc.; 2012. p 2951-2959.

75. Cordova JS, Gurbani SS, Holder CA, Olson JJ, Schreibmann E, Shi R, Guo Y, Shu H-KG, Shim H, Hadjipanayis CG. Semi-Automated Volumetric and Morphological Assessment of Glioblastoma Resection with Fluorescence-Guided Surgery. Mol Imaging Biol 2016;18(3):454-462.

76. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015.

77. Bishop C. Pattern Recognition and Machine Learning. Jordan MIN, Robert; Schoelkopf, Bernhard, editor. New York, NY: Springer-Verlag; 2006. 738 p.

78. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representaitons. San Diego, CA, USA2015.

79. Bergstra JS, Bardenet R, Bengio Y, Kégl B. Algorithms for Hyper-Parameter Optimization. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors. Adv Neur In: Curran Associates, Inc.; 2011. p 2546-2554.

80. Jones DR. A Taxonomy of Global Optimization Methods Based on Response Surfaces. J Global Optimiz 2001;21(4):345-383.

81. Selaraju R, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization. CoRR 2016;abs/1610.02391.

82. Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27(8):861-874.

83. García-Gómez JM, Luts J, Julià-Sapé M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-García E, Olier I, Postma G, Monleón D, Moreno-Torres À, Pujol J, Candiota A-P, Martínez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arús C, Robles M. Multiproject–multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. Magn Reson Mater Phy 2008;22(1):5.

84. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. arXiv preprint arXiv:171005381 2017.

85. Wen PY, Macdonald DR, Reardon DA. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. Journal of Clinical Oncology 2010.

86. Stupp R, Hegi ME, Gilbert MR. Chemoradiotherapy in malignant glioma: standard of care and future directions. Journal of Clinical Oncology 2007.

87. Stupp R, Hegi ME, Mason WP, van den Bent MJ. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year …. The lancet oncology 2009.

88. Stupp R, Mason WP, Bent VMJ. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. … England Journal of … 2005.

89. Tsuchiya K, Mizutani Y. Preliminary evaluation of fluid-attenuated inversion-recovery MR in the diagnosis of intracranial tumors. American Journal of … 1996.

90. Stupp R, Hegi ME, Gilbert MR, Chakravarti A. Chemoradiotherapy in malignant glioma: standard of care and future directions. J Clin Oncol 2007;25(26):4127-4136.

91. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff R-O, European Organisation for R, Treatment of Cancer Brain T, Radiation Oncology G, National Cancer Institute of Canada Clinical Trials G. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol 2009;10(5):459-466.

92. Stupp R, Taillibert S, Kanner AA, et al. Maintenance therapy with tumor-treating fields plus temozolomide vs temozolomide alone for glioblastoma: A randomized clinical trial. JAMA 2015;314(23):2535-2543.

93. Fitzek MM, Thornton AF, Rabinov JD, Lev MH, Pardo FS, Munzenrider JE, Okunieff P, Bussière M, Braun I, Hochberg FH, Hedley-Whyte ET, Liebsch NJ, Harsh GR. Accelerated fractionated proton/photon irradiation to 90 cobalt gray

equivalent for glioblastoma multiforme: results of a phase II prospective trial. Journal of Neurosurgery 1999;91(2):251-260.

94. Maudsley AA, Domenig C, Sheriff S. Reproducibility of serial whole-brain MR Spectroscopic Imaging. NMR in Biomedicine 2010.

95. Cordova JS, Kandula S, Gurbani S, Zhong J, Tejani M, Kayode O, Patel K, Prabhu R, Schreibmann E, Crocker I, Holder CA, Shim H, Shu H-K. Simulating the Effect of Spectroscopic MRI as a Metric for Radiation Therapy Planning in Patients with Glioblastoma. Tomography : a journal for imaging research 2016;2(4):366-373.

96. Gurbani SS, Schreibmann E, Sheriff S, Cooper LAD, Shu HKG, Holder CA, Maudsley AA, Shim H. A Software Platform for Collaborative Radiation Therapy Planning Using Spectroscopic MRI. International Journal of Radiation Oncology • Biology • Physics 2017;99(2):E667.

97. Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. Journal of Open Source Software 2016;1:26.

98. Yoo TS, Ackerman MJ, Lorensen WE, Schroeder W, Chalana V, Aylward S, Metaxas D, Whitaker R. Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit. Studies in health technology and informatics 2002:586-592.

99. Engwer C, Hillen T, Knappitsch M, Surulescu C. Glioma follow white matter tracts: a multiscale DTI-based model. Journal of mathematical biology 2015;71(3):551-582.

100. Gurbani SS, Schreibmann E, Sheriff S, Holder CA, Cooper LAD, Maudsley A, Shim H. Rapid Internal Normalization of Spectroscopic MRI Maps Using a Gaussian Mixture Model. 2017; Denver, CO. p TU-AB-601-610.

101. Maudsley AA, Domenig C. Signal normalization for MR spectroscopic imaging using an interleaved water-reference. International Society for Magnetic … 2008.

102. Keles GE, Chang EF, Lamborn KR, Tihan T, Chang C-J, Chang SM, Berger MS. Volumetric extent of resection and residual contrast enhancement on initial surgery as predictors of outcome in adult patients with hemispheric anaplastic astrocytoma. Journal of Neurosurgery 2006;105(1):34-40.

103. Kurita T, Otsu N. Texture Classification by Higher Order Local Autocorrelation. 1993. p 175-178.

104. Cordova JS, Schreibmann E, Hadjipanayis CG, Guo Y, Shu H-KG, Shim H, Holder CA. Quantitative Tumor Segmentation for Evaluation of Extent of Glioblastoma Resection to Facilitate Multisite Clinical Trials. Translational oncology 2014;7(1):40.

105. RTOG. Randomized Phase II Trial of Hypofractionated Dose-Escalated Photon IMRT or Proton Beam Therapy Versus Conventional Photon Irradiation With Concomitant and Adjuvant Temozolomide in Patients With Newly Diagnosed Glioblastoma. Radiat Ther Oncol Gr NRG-BN001 Protoc Inf 2014.

106. Gurbani SS, Schreibmann E, Maudsley AA, Cordova JS, Soher BJ, Poptani H, Verma G, Barker PB, Shim H, Cooper LAD. A convolutional neural network to filter artifacts in spectroscopic MRI. Magnetic Resonance in Medicine 2018.

107. Gurbani SS, Sheriff S, Maudsley AA, Shim H, Cooper LAD. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. Magnetic resonance in medicine 2019.

108. Stadlbauer A, Buchfelder M, Doelken MT, Hammen T, Ganslandt O. Magnetic resonance spectroscopic imaging for visualization of the infiltration zone of glioma. Central European Neurosurgery 2011;72(02):63-69.

109. Chronaiou I, Stensjøen AL, Sjøbakk TE, Esmaeili M, Bathen TF. Impacts of MR spectroscopic imaging on glioma patient management. Acta Oncologica 2014;53(5):580-589.

110. Ken S, Vieillevigne L, Franceries X, Simon L, Supper C, Lotterie J-A, Filleron T, Lubrano V, Berry I, Cassol E. Integration method of 3D MR spectroscopy into treatment planning system for glioblastoma IMRT dose painting with integrated simultaneous boost. Radiation Oncology 2013;8(1):1.

111. Nelson SJ, Graves E, Pirzkall A, Li X, Antiniw Chan A, Vigneron DB, McKnight TR. In vivo molecular imaging for planning radiation therapy of gliomas: an application of 1H MRSI. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 2002;16(4):464-476.

112. Pirzkall A, McKnight TR, Graves EE, Carol MP, Sneed PK, Wara WW, Nelson SJ, Verhey LJ, Larson DA. MR-spectroscopy guided target delineation for high-grade gliomas. International Journal of Radiation Oncology* Biology* Physics 2001;50(4):915-928.

113. Gurbani S, Mellon E, Weinberg B, Schreibmann E, Maudsley AA, Sheriff S, Barker PB, Kleinberg L, Cooper LAD, Shu H-K, Shim H. A Feasibility Study of Radiation Therapy Dose Escalation Guided by Spectroscopic Magnetic Resonance Imaging in Patients with Glioblastoma. 2018; Paris, FR. p Abstract #1155.

114. Mierisová Š, Ala-Korpela M. MR spectroscopy quantitation: a review of frequency domain methods. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo 2001;14(4):247-259.

115. Vanhamme L, Sundin T, Hecke PV, Huffel SV. MR spectroscopy quantitation: a review of time-domain methods. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo 2001;14(4):233-246.

116. Stefan D, Di Cesare F, Andrasescu A, Popa E, Lazariev A, Vescovo E, Strbak O, Williams S, Starcuk Z, Cabanas M. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. Measurement Science and Technology 2009;20(10):104035.

117. Reynolds G, Wilson M, Peet A, Arvanitis TN. An algorithm for the automated quantitation of metabolites in in vitro NMR signals. Magnetic Resonance in Medicine 2006;56(6):1211-1219.

118. Kyathanahally SP, Mocioiu V, Pedrosa de Barros N, Slotboom J, Wright AJ, Julià-Sapé M, Arús C, Kreis R. Quality of clinical brain tumor MR spectra judged by humans and machine learning tools. Magnetic resonance in medicine 2018;79(5):2500-2510.

119. Gurbani SS, Schreibmann E, Maudsley AA, Cordova JS, Soher BJ, Poptani H, Verma G, Barker PB, Shim H, Cooper LAD. A convolutional neural network to filter artifacts in spectroscopic MRI. Magnetic Resonance in Medicine 2018;80(5):1765-1775.

120. Liou C-Y, Huang J-C, Yang W-C. Modeling word perception using the Elman network. Neurocomputing 2008;71(16):3150-3157.

121. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504-507.

122. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 2002;47(6):1202-1210.

123. Goodfellow I, Bengio Y, Courville A. Deep Learning: MIT Press; 2016. 800 p.

124. Donoho DL, Johnstone IM. Adapting to Unknown Smoothness via Wavelet Shrinkage. J Am Stat Assoc 1995;90(432):1200-1224.

125. Daubechies I. Ten lectures on wavelets: Siam; 1992.

126. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M. Tensorflow: a system for large-scale machine learning. 2016. p 265-283.

127. Strang G, Nguyen T. Wavelets and filter banks: SIAM; 1996.

128. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436-444.

129. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26(3):297-302.

130. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, Wells WM, Jolesz FA, Kikinis R. Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index: Scientific Reports. Academic radiology 2004;11(2):178-189.

131. Vo K, Pham D, Nguyen M, Mai T, Quan T. Combination of Domain Knowledge and Deep Learning for Sentiment Analysis. 2017. Springer. p 162-173.

132. Yu T, Jan T, Simoff S, Debenham J. Incorporating prior domain knowledge into inductive machine learning.

133. Heuer A. A new algorithm for automatic phase correction by symmetrizing lines. Journal of Magnetic Resonance (1969) 1991;91(2):241-253.